

# Simulación de reacciones en biomoléculas con QM/MM



**Jesús I. Mendieta-Moreno**

Department of Física Teórica de la Materia Condensada  
Universidad Autónoma de Madrid

Esta tesis se presenta para el grado de  
*Doctor en Ciencias Físicas*

Facultad de Ciencias

May 2017





## RESUMEN

---

En esta tesis se desarrollara un método de simulación híbrido QM/MM que permite simular biomoléculas con una región de los átomo en cuántica (QM) y el resto del sistema con un potencial clásico (MM). El cálculo cuántico se llevara a cabo con el programa FIREBALL que utiliza la teoría del funcional de la densidad (DFT) para simular la estructura electrónica. FIREBALL tiene la ventaja de ser un código DFT de orbitales localizados que trabaja en espacio real y tiene una gran eficiencia computacional manteniendo una buena precisión. Gracias a esto es un método que se ajusta perfectamente al tipo de simulación necesaria para hacer estudios de procesos biológicos. Para la región clásica se utilizará el código AMBER que permite simular una gran cantidad de átomos de forma eficiente con un campo de fuerzas empírico, además tiene multitud de funciones que permiten explorar el espacio conformacional deseado. En el desarrollo de este método se prestara especial atención en como los átomos de la región clásica son capaces de modificar la estructura electrónica a través de las cargas parciales del campo de fuerza de MM. También se estudiará el efecto de esta modificación de la estructura electrónica sobre un puente salino formado por la interacción de un Aspártico y una Lisina comprobando que el programa reproduce los efectos conocidos de como el ambiente afecta a la estructura electrónica.

El método DFT FIREBALL utiliza una base de orbitales localizados para conseguir una gran balance entre eficiencia computacional y precisión.

Para conseguir este objetivo se desarrollará una base lo mas eficiente posible computacionalmente y se optimizara para el estudio de sistema biológicos. Los elementos que formaran la base son hidrógeno, carbono, nitrógeno, oxígeno, azufre, fósforo y magnesio. Estos elementos son de los más abundantes en biología y fundamentales en la mayoría de reacciones enzimáticas.

A continuación se aplicará este método QM/MM FIREBALL/AMBER para la simulación de reacciones enzimáticas teniendo en cuenta la dinámica de estos procesos. Para estudiar reacciones de desarrollara una metodología para explorar el espacio conformacional deseado de la reacción de forma dinámica y poder calcular la energía libre a lo largo de una superficie 2D y la trayectorias de reacción posibles. La primera reacción enzimática que se estudiara es la escisión de la molécula de ARN catalizada por la proteína RNasa A. En esta reacción se observaran dos caminos posibles sobre la superficie de energía y se estudiara cual es mas probable enérgicamente. También se observaran los diferentes que hay a lo largo de las trayectorias de la reacción y que residuos hacen posible catalizar la reacción disminuyendo la barrera de energía. Posteriormente se simulara la reacción de inter-conversión reversible entre dihidroxiacetona fosfato (DHAP)y gliceraldehido 3-fosfato (GAP) catalizada por la proteína Triosa-fosfato isomerasa (TIM). Para esta enzima se calculará también una superficie de energía libre y se determinara la trayectoria mas probable de la reacción. Para esta reacción se propondrá un nuevo camino de reacción gracias al estudio dinámico de las reacciones.

Por último se estudiara la reacción foto-química de formación del dímero de ciclobutano de timina (CTD) en el ADN. Para estudiar la reacción se harán mapas adiabáticos de energía libre del estado fundamental y el primer estado excitado y se determinara que región es susceptible de ir a la intersección cónica una vez se produce la excitación al absorber un fotón. También se estudiará con dinámicas no adiabáticas que sucede a

las trayectorias que llegan a la intersección cónica donde se produce la desexcitación y pueden formar el CTD o volver a la forma canónica del ADN. Tanto en los mapas adiabáticos como en las trayectorias no adiabáticas se vea como la estructura de ADN previene la formación del dímero de timina protegiéndonos de mutaciones en el ADN.



# ÍNDICE GENERAL

---

|  |            |
|--|------------|
| <b>Resumen</b>   | <b>III</b> |
| <b>1. Introduction</b>   | <b>1</b>   |
| 1.1. Dinámica de proteínas y ácidos nucleicos . . . . .                                      | 2          |
| 1.2. Dinámica molecular en biología . . . . .  | 8          |
| 1.2.1. Métodos QM/MM . . . . .   | 9          |
| <b>2. Fireball/Amber</b>   | <b>11</b>  |
| 2.1. Introducción . . . . .  | 11         |
| 2.2. Descripción del Método . . . . .  | 13         |
| 2.2.1. Método QM: Introducción . . . . .   | 14         |
| 2.2.2. Método QM: DFT . . . . .  | 15         |
| 2.2.3. Método QM: FIREBALL . . . . .   | 19         |
| 2.2.4. FIREBALL / AMBER . . . . .  | 22         |
| 2.3. Ejemplo: efecto de las cargas de MM en la región de QM en<br>un puente salino . . . . . | 30         |
| <b>3. Optimización de orbitales FIREBALL</b>   | <b>35</b>  |
| 3.1. Method . . . . .  | 36         |
| 3.1.1. Orbitales numéricos tipo atómico . . . . .  | 36         |
| 3.1.2. Optimización: Bases de datos y benchmarks . . . . .                                   | 39         |
| 3.2. Resultados . . . . .  | 40         |
| 3.2.1. C, H, O, N . . . . .  | 40         |

|   |            |
|---|------------|
| 3.2.2. S, P, Mg . . . . .   | 44         |
| 3.3. Test in QM/MM simulations . . . . .                          | 48         |
| 3.4. Conclusiones . . . . .                                       | 49         |
| <b>4. Estudio de reacciones enzimáticas</b>                       | <b>51</b>  |
| 4.1. Enzimas . . . . .  | 51         |
| 4.1.1. Unión del sustrato . . . . .                               | 52         |
| 4.1.2. Catálisis . . . . .  | 54         |
| 4.1.3. Salida del producto . . . . .                              | 59         |
| 4.2. Estudio dinámico de reacciones en biomoléculas . . . . .     | 59         |
| 4.3. RNasa A . . . . .  | 64         |
| 4.3.1. Introducción . . . . .                                     | 64         |
| 4.3.2. Resultados . . . . .                                       | 64         |
| 4.3.3. Discusión . . . . .  | 71         |
| 4.4. Triosa-fosfato isomerasa . . . . .                           | 72         |
| 4.4.1. Introducción . . . . .                                     | 72         |
| 4.4.2. Resultados . . . . .                                       | 74         |
| 4.4.3. Discusión . . . . .  | 80         |
| 4.5. Conclusión . . . . .   | 82         |
| <b>5. Foto-reacciones en ADN</b>                                  | <b>85</b>  |
| 5.1. Dinámica molecular con transiciones electrónicas . . . . .   | 86         |
| 5.2. Simulación de una reacción fotoquímica en ADN: CTD . . . . . | 90         |
| 5.2.1. Introducción . . . . .                                     | 90         |
| 5.2.2. Método . . . . .   | 92         |
| 5.2.3. Resultados: estudio adiabático . . . . .                   | 96         |
| 5.2.4. Resultados: estudio no adiabático . . . . .                | 103        |
| 5.2.5. Discusión y conclusiones . . . . .                         | 107        |
| <b>6. Conclusiones generales</b>                                  | <b>113</b> |

|   |            |
|---|------------|
| Índice general  | IX         |
| <b>Apéndice Tablas y gráficas de la optimización de la base</b> | <b>119</b> |
| <b>Bibliografía</b>   | <b>129</b> |





# CAPÍTULO 1

## INTRODUCCIÓN

---

Al intentar entender de forma mas profunda el funcionamiento de los procesos biológicos se están utilizando cada vez mas herramientas distintas que nos ayudan a descifrar que pasa dentro de un ser vivo. Estas herramientas cada vez tienen una frontera menos clara entre la biología, la química y la física ya que cada una aporta a aumentar el conocimiento sobre estos procesos. En concreto la física está aportando a la biología nuevos métodos para poder entender los sistemas tanto experimentalmente con modelos teóricos y computacionales.

Las técnicas experimentales de la física han permitido grandes avances en biología, algunos de ellos estructurales gracias a cristalografía y la resonancia magnética nuclear que han permitido la resolución atómica de grandes biomoléculas. Y también se han podido entender procesos dinámicos gracias a técnicas físicas como la fluorescencia o la espectroscopia molecular.

Debido a que las leyes de la física son universales se pueden aplicar a cualquier fenómeno natural como son los procesos biológicos otro gran aporte que puede hacer la física es utilizar las herramientas matemáticas se han desarrollado para modelar estos sistemas. Los modelos nos permiten estudiar de forma mas precisa los sistemas biológicos y hacer variaciones

sobre los modelos para al compararlos con los experimentos intentar comprender que pasa dentro de una célula.

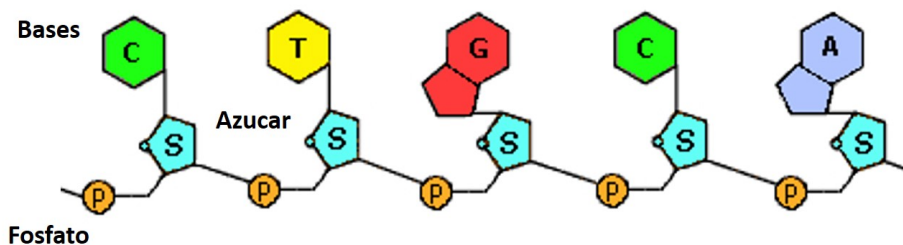
La biología molecular tiene un enfoque en el cual se intenta explicar los fenómenos que pasan en un ser vivo como una serie de interacciones entre moléculas que al sumirlas todas forman un ser vivo. Este enfoque es muy parecido a la termodinámica y a la física de sistemas complejos donde una gran cantidad de interacciones forman un sistema mucho mas complejo y aparecen propiedades emergentes. Estos sistemas además de tener muchas partículas suelen estar a temperatura ambiente por lo que las herramientas de la mecánica estadística nos serán de gran ayuda para modelizar los sistemas.

Un método habitual para estudiar los sistemas termodinámicos es hacer simulaciones de dinámica molecular en los cuales se pueden tener muchas partículas y ver como afecta de forma estadísticas las distintas interacciones que tienen a variables como la temperatura o la presión. Además al hacer dinámicas se pueden calcular ciertas variables como la variación en energía libre de un proceso que se pueden trasladar a mediciones experimentales.

Esta modelización también se puede hacer para distintas aproximaciones desde considerar cada molécula como una única partícula o tener en cuenta todos los átomos o incluso los electrones del sistema. Este tipo de variación en la aproximación nos permitirá diferenciar cual es la parte importante en cada proceso y avanzar hacia un conocimiento mas completo de la biología.

## **1.1. Dinámica de proteínas y ácidos nucleicos**

En esta tesis se estudiara computacionalmente las biomoléculas mas grande que hay en las células, que son las proteínas y los ácidos nucleicos.

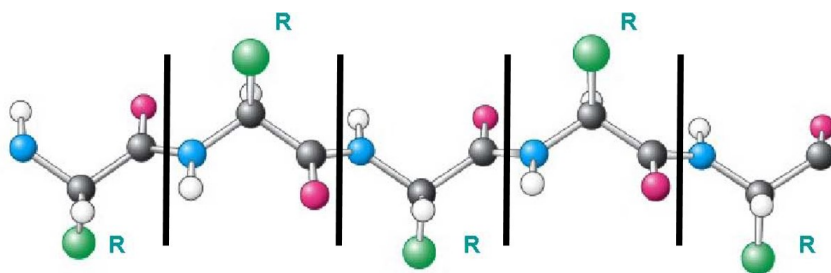


**Figura 1.1** Esquema de la cadena de ADN

Los ácidos nucleicos ADN y ARN son biomoléculas esenciales para cualquier ser vivo. El ADN funciona como banco de dato donde tiene toda la información para que la célula pueda funcionar. EL ARN está involucrado en la síntesis de proteína (tARN), en la construcción del ribosoma (rARN), en transportar la información genética (mARN) y en mecanismos de regulación. Su estructura está formada por una cadena de fosfatos y azúcares (desoxirribosa en el ADN y ribosa en el ARN) con una base distinta saliendo de cada azúcar, ver figura 1.1.

Estructuralmente el ADN suele estar en cadena doble formando una doble hélice<sup>1</sup>. Esta estructura de doble hélice se forma gracias a los puentes de Watson y Crick donde cada base tiene en frente a su complementaria. Esta estructura es una de las más estables que existen en biomoléculas. Esta doble hélice puede cambiar un poco estructuralmente en función del ambiente. El ARN en cambio suele estar en cadena simple y puede tener una estructura tridimensional fija que depende de su secuencia como en el caso de los ARN transferente. Además al estar normalmente en cadena simple es mucho más móvil.

Las proteínas son biomoléculas que están envueltas en todos los pasos de la actividad celular y pueden tener diferentes funciones: reguladoras, estructurales, catalíticas, motoras, etc. Las proteínas son heteropolímeros formadas por una cadena lineal de distintos aminoácidos y unidos entre sí por un enlace peptídico. Los aminoácidos son biomoléculas pequeñas que



**Figura 1.2** Imagen de los átomos de la cadena principal de una proteína. Con R esta marcado el átomo donde comienza la cadena secundaria de cada aminoácido.

tienen una parte común que formará la cadena principal de la proteína al irse uniendo para formar el heteropolímero, ver figura 1.2. Por otro lado cada tipo de los 20 aminoácidos distintos tiene una cadena lateral específica que varía en tamaño, forma, carga, hidrofobicidad y reactividad. Cada proteína esta definida por su estructura primaria que es la secuencia de los aminoácidos que la componen.

Las proteínas al formarse juntando aminoácidos habitualmente forman una estructura tridimensional única que viene dada por la secuencia. Esta estructura tridimensional es un mínimo de energía libre que se conoce como estructura plegada de la proteína y está estabilizada por una gran cantidad de distintas interacciones que contribuyen en la estabilización energética de la estructura.

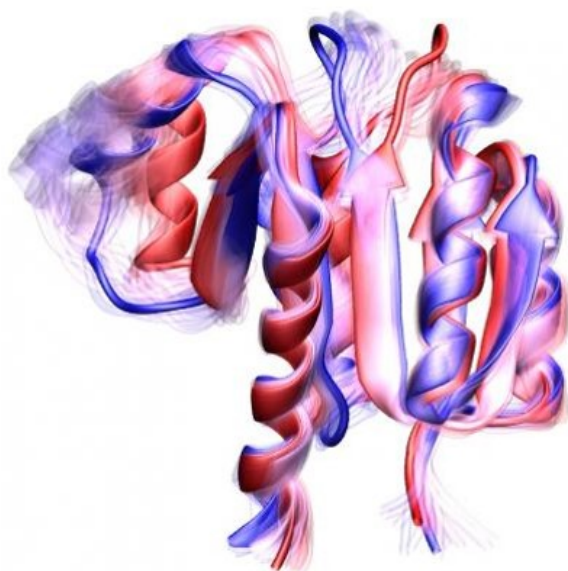
Los estudios sobre las estructura han sido típicamente estáticos y han revelado mucha información sobre la actividad tanto de los ácidos nucleicos como de las proteínas. Sin embargo sabemos que la vida es un fenómeno que está en constante movimiento. EL movimiento en los seres vivos lo podemos ver en todas las escalas posibles que van desde el movimiento de un animal a las vibraciones a escala atómica. A nivel microscópico, dentro de una célula, se puede observar este movimiento en mecanismo como la división celular que implican muchísimas biomoléculas de todos los tamaños moviéndose de forma

orquestrada para poder llevar a cabo la división. También se puede ver el movimiento en escalas mas pequeñas como en las proteínas motoras, por ejemplo quinesina que es capaz de avanzar por un filamento al transformar energía química del ATP en energía mecánica.

Estos procesos en la célula ocurren a escalas temporales muy distintas que pueden llegar a 15 ordenes de magnitud de diferencia. Estas escalas temporales están relacionadas con la escala espacial a la que suceden los cambios. Esto se debe a que los cambios tienen que ver con una variación en la función de las moléculas y de sus interacciones, ya sea entre moléculas distintas o internamente en la molécula. Los distintos eventos bioquímicos que sucedan en las biomoléculas se van a poder describir por cantidades físicas computables y estas dependerán de la escala temporal en la que sucedan.

Para los estudios que se verán posteriormente nos importa que sucede a nivel intra-molecular, donde las cantidades físicas las podremos describir como un conjunto de estructuras a nivel atómico. Que una medida física dependa de un conjunto se debe a que todos los átomos del sistema con un movimiento térmico continuo que intercambia energía con el medio a través de colisiones de partículas. En el caso de una proteína la estructura plegada no es una estructura donde los átomos estén en unas coordenadas definidas sino que es un conjunto de estructuras que fluctúan debido a la temperatura, ver figura 1.3. Las estructuras que se visitan dependen de la energía del sistema, en condiciones normales (298 K y 1 atm) la media de energía térmica  $k_B T$  es de  $0.593 \text{ Kcal/mol}$ . Como la temperatura sigue una distribución de Maxwell de las velocidades puede haber fluctuaciones mayores y visitar estructuras con energías mas altas aunque la probabilidad de visitarlas ira disminuyendo de forma exponencial.

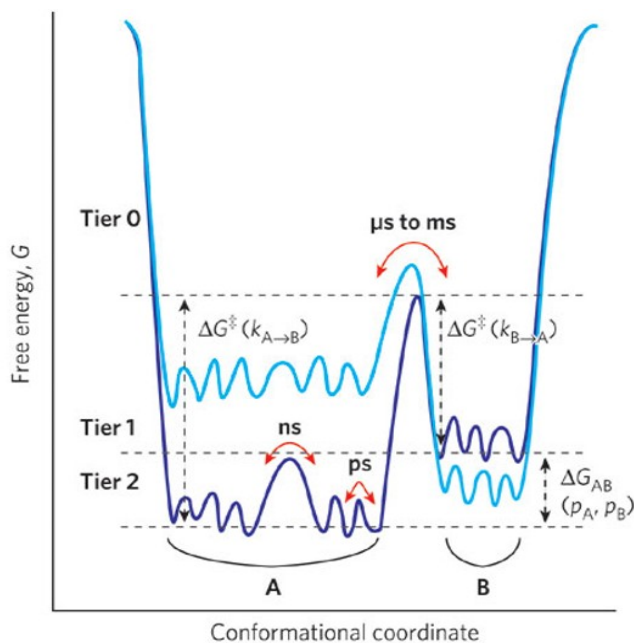
La escala temporal de los movimientos intra-moleculares va a depender de la energía ya que como hemos visto antes la probabilidad de visitar un estado depende de ella. Por lo que si la energía es muy alta sera mas



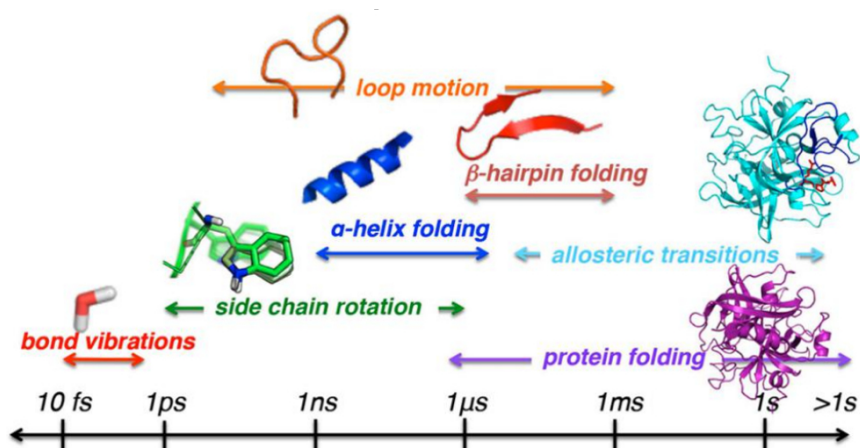
**Figura 1.3** Esquema del conjunto de estructuras que tiene una proteína en su estado plegado.

improbable que este estado se visite pero si paso suficiente tiempo se visitara. Como los diferentes estados en las biomoléculas suelen tener una barrera energética el tiempo para pasar a otro estado dependerá de esta barrera, ver 1.4.

Estás variaciones en la escala temporal de los movimientos debido a la energía de la barrera hace que exista una especie de jerarquía en los movimientos. En la cual los movimientos mas rápidos serán los que prácticamente no tengan barrera o esta este dentro de la energía térmica. Entre estos movimientos suelen estar la vibración de los enlaces (entre  $\sim 10$  fs y  $\sim 1$  ps) y los movimientos de las cadenas laterales(entre  $\sim 1$  ps y  $\sim 1$  ns). Ya en movimientos mas grandes donde hay que mover además de la cadena lateral de varios residuos la principal el tiempo suele ser cada vez mayor y llegar a tardar microsegundos o segundos.



**Figura 1.4** Esquema de la dependencia temporal para pasar de un estado a otro en función de la barrera de energía.



**Figura 1.5** Esquema de la escala temporal de los movimientos intramoleculares y de su jerarquía.

## 1.2. Dinámica molecular en biología

Como las cantidades físicas medible de las biomoléculas son conjuntos de estructuras si queremos medir esto computacionalmente necesitaremos un método que nos permita obtener ese conjunto. Estos sistemas al tener muchas partículas interactuando no se pueden resolver analíticamente. Para resolver este problema a finales de 1950 se elaboro un método donde las ecuaciones del movimiento de Newton se resolvían de forma numérica y las fuerzas eran calculadas a partir de la energía potencial de interacción entre las partículas.<sup>2-4</sup> Este método fue aplicado en física teórica para estudiar distintos sistemas. Este método se fue extendiendo y utilizándose cada vez mas para sistemas físicos.

Aunque no fue hasta 1977 cuando la dinámica molecular se uso por primera vez en biología para simular el inhibidor de la tripsina pancreática bovina por Mccammon et al.<sup>5</sup>. En el cual se simulo una por primera vez una proteína y aunque fue en vacío y solo 9,2 ps ya sugerían que se comportaba como una estructura dinámica. Posteriormente las dinámicas moleculares se extendieron en biología simulando proteínas y ácidos nucleicos, ayudando a entender los distintos aspectos que afectan a a estructura y dinámica de las biomoléculas. Así como también ayudaran a entender de forma mas precisa algunos experimentos. Con ello mejoro la calidad de los campos de fuerza y los programas de simulación como CHARMM<sup>?</sup>, AMBER<sup>?</sup>, GROMACS<sup>?</sup>, NAMD<sup>?</sup>, haciendo cada vez mas accesible los estudios de biomoléculas de forma computacional. Este aumento fue sobre todo a la capacidad que tiene la dinámica de poder explicar que pasa a escala atómica y en la escala temporal.

La gran ventaja de la dinámica molecular para estudiar sistemas biológicos es que como hemos comentado antes las biomoléculas son sistemas que están en movimiento de forma permanente y los estados físicos hay que definirlos como un conjunto de estructuras. Los programas



de dinámica molecular al trabajar con trayectorias y a una temperatura son capaces de tener estos conjuntos de estructuras en cuenta y trabajar con energía libre. Además en las simulaciones se pueden controlar parámetros o hacer mutaciones en residuos y ver que efecto tienen en la dinámica de estas estructuras.

En los últimos años la dinámica molecular en biología ha avanzado muchísimo y ha utilizado en muchas escalas. Tanto para estudiar difusión de proteínas en fluido y como se estructuran a estudios de comportamientos cuánticos en proteínas, aunque las simulaciones mas utilizadas siguen siendo las de escala atómica.

### 1.2.1. Métodos QM/MM

Una de los métodos más utilizados para estudiar reacciones en biomoléculas es el método híbrido QM/MM. Estos métodos están basados en poder definir dos regiones de la biomolécula de forma distinta, una mas pequeña se utilizara un método cuántico (QM) en el que se tendrá en cuenta la estructura electrónica y el resto del sistema estará modelado por un potencial empírico de mecánica clásica (MM). Este tipo de método fue introducido por Warshel y Levitt en 1976<sup>6</sup> en el cual introducían una energía debido al ambiente proteico a los cálculos de energía potencial de QM para estructuras estáticas. Aunque no fue hasta 1990 cuando se empezó a utilizar de forma mas extensa después del trabajo de Field, Bash y Karplus<sup>7</sup> en el que se propuso el esquema que se utiliza actualmente. A partir de ese momento los programas de dinámica molecular empezaron a implementar estos métodos híbridos en sus paquetes de programas y se popularizo el uso para estudiar reacciones enzimáticas. En la mayoría de los estudios de reacciones debido al coste computacional de la región de QM no se utiliza dinámica molecular y se buscan solo los estados estables y los estados de transición<sup>8</sup>. En los últimos años esto está cambiando y en esta tesis presentamos un método con una relación eficiencia computacional

y precisión que permite estudiar las reacciones enzimáticas de forma dinámica.

## CAPÍTULO 2

# UN MÉTODO QM/MM PARA LA SIMULACIÓN DE BIOMOLÉCULAS: FIREBALL/AMBER

---

### 2.1. Introducción

Uno de los retos más interesantes que se presentan ahora mismo en la Biología Molecular y Biofísica es entender cómo funcionan las proteínas desde un punto de vista atómico. Algunos procesos como cambios conformacionales<sup>9</sup> o cómo se une un ligando a una proteína<sup>10</sup> se pueden estudiar con simulaciones de Mecánica Molecular (MM) utilizando campos de fuerza empíricos.

En cambio para estudiar reacciones biomoleculares donde los enlaces químicos se tienen que romper o formar, los campos de fuerza clásicos no son capaces de describirlos con suficiente exactitud. Esto se debe a que la naturaleza de estos enlaces es dependiente de la estructura electrónica y para poder describirlos con exactitud hay que hacer cálculos de Mecánica Cuántica (QM).<sup>11,12</sup>

Por desgracia el coste computacional de estos métodos es muy grande pero para solucionar esto se puede utilizar una aproximación QM/MM donde una región se calcula con mecánica cuántica, típicamente el centro activo, mientras que el resto del sistema se describe con Mecánica

Molecular. Esta aproximación nos reduce el coste computacional pero sigue manteniendo una precisión suficiente para describir los procesos que tienen lugar en el centro activo.<sup>6-8</sup>

La precisión y el coste computacional de un cálculo QM/MM depende mucho del método de cuántica que se utilice. En la mayoría de las aplicaciones en biomoléculas se necesita estudiar el espacio conformacional de la reacción de forma detallada para poder entender el mecanismo catalítico. Si se quiere estudiar un gran espacio conformacional se necesitan métodos de QM muy eficientes como son los métodos semi-empíricos<sup>13</sup>, sin embargo la precisión que tienen estos métodos puede no ser suficiente. Por otro lado podríamos utilizar aproximaciones de alta precisión en química cuántica pero estas suelen tener un coste computacional prohibitivo y nos reduce la cantidad de espacio conformacional que podemos estudiar.<sup>14</sup>

Para buscar un buen balance entre eficiencia computacional y precisión hemos desarrollado una nueva implementación de QM/MM<sup>15</sup> basada en la combinación de FIREBALL<sup>16-20</sup> para la región de QM y AMBER<sup>21,22</sup> para la región clásica (MM). FIREBALL es un programa basado en la teoría del funcional de la densidad (DFT) muy adecuado para el estudio de procesos biomoleculares con QM/MM ya que debido a su eficiencia computacional nos permite poder trabajar haciendo simulaciones de Dinámica Molecular (MD) con una región de QM con bastantes átomos manteniendo una precisión bastante elevada.<sup>15</sup> La región que se modela con mecánica molecular se hará con el software de dinámica molecular AMBER que tiene una gran reputación y además ofrece una plataforma con muchas herramientas para el estudio de simulaciones MD. En concreto ofrece una gran cantidad de aplicaciones para poder estudiar un gran espacio conformacional (umbrella sampling, replica exchange, nudged elastic band, targeted MD, steered MD, ...) <sup>21</sup> y las podemos aplicar directamente con QM/MM.

## 2.2. Descripción del Método

Para realizar simulaciones de QM/MM MD necesitamos resolver la estructura electrónica del sistema en cada paso de dinámica ( $t_i = m_i \Delta t$ ;  $\Delta t \sim 10^{-15}$  s) de la simulación. Los métodos mas prácticos para hacer simulaciones de primeros principios están basados en la teoría del funcional de la densidad (DFT)<sup>23,24</sup> ya que ofrece una simplificación importante del problema de muchos cuerpos.<sup>11</sup> FIREBALL<sup>16-20</sup> es un método de DFT diseñado para poder estudiar sistemas complejos usando técnicas de dinámica molecular. En FIREBALL la estructura electrónica se calcula resolviendo las ecuaciones de DFT (ver posteriormente) usando una base de orbitales localizados de tipo atómico en los átomos del sistema. Los orbitales utilizados son de corto alcance, numéricos y están optimizados de forma que se pueda obtener una buena precisión con pocos orbitales por átomo.<sup>20,25</sup>

La gran mayoría de los métodos DFT usan condiciones de límite periódicas (ondas planas, representación de la densidad electrónica en un reticulado periódico, uso de transformadas de fourier rápidas, etc) ya que estas técnicas fueron desarrolladas en física del estado sólido para estudiar sólidos en estructura cristalina donde los átomos están ordenados de forma periódica. En cambio en los sistemas biomoleculares los átomos no suelen tener una estructura periódica y por lo tanto no existe ninguna ventaja al utilizar una periodicidad artificial en los cálculos. FIREBALL es una técnica de DFT que está completamente formulada en *espacio real* y por eso es innecesario definir superceldas para hacer un cálculo. Esta propiedad hace que *Fireball* pueda trabajar de forma muy efectiva para realizar simulaciones en biomoléculas en combinación con cálculos de MM (que también son métodos de espacio real).

### 2.2.1. Método QM: Introducción

Para poder resolver de forma más detallada la energía y la fuerza de un sistema es necesario calcular la estructura electrónica que tiene. Para poder calcular esta estructura electrónica hay que resolver la ecuación de Schrödinger. Para poder resolver este tipo de sistemas lo primero que hacemos es separar el hamiltoniano entre los núcleos  $\{\vec{R}_\alpha, M_\alpha\}$  y electrones  $\{\vec{r}_i, m\}$ , y tendríamos el siguiente hamiltoniano:

$$\hat{H} = -\sum_{\alpha} \frac{p_{\alpha}^2}{2M_{\alpha}} - \sum_i \frac{p_i^2}{2m} + \frac{1}{2} \sum_{i,j} \frac{1}{|\vec{r}_i - \vec{r}_j|} - \sum_{i,\alpha} \frac{Z_{\alpha}}{|\vec{r}_i - \vec{R}_{\alpha}|} + \frac{1}{2} \sum_{\alpha,\beta} \frac{Z_{\alpha}Z_{\beta}}{|\vec{R}_{\alpha} - \vec{R}_{\beta}|} \quad (2.1)$$

Si ahora utilizamos la aproximación de Born-Oppenheimer, donde el movimiento de los electrones y los núcleos se desacoplan, podemos ver la energía total del sistema como la energía cinética de los núcleos mas una energía 'potencial'  $\mathcal{E}(\{\vec{R}_{\alpha}\})$  de la nube electrónica que es función de las coordenadas de los átomos.

$$E_{tot} = \sum_{\alpha} \frac{p_{\alpha}^2}{2M_{\alpha}} + \mathcal{E}(\{\vec{R}_{\alpha}\}) \quad (2.2)$$

Donde  $\mathcal{E}(\{\vec{R}_{\alpha}\})$  es la energía del estado basal del sistema en una configuración fija de los núcleos  $\{\vec{R}_{\alpha}\}$ ,

$$\left( \hat{H} - \sum_{\alpha} \frac{p_{\alpha}^2}{2M_{\alpha}} \right) \Psi_{\{\vec{R}_{\alpha}\}}\{\vec{r}_i\} = \mathcal{E}(\{\vec{R}_{\alpha}\}) \Psi_{\{\vec{R}_{\alpha}\}}\{\vec{r}_i\} \quad (2.3)$$

Y podemos calcular la fuerza que se ejerce sobre los núcleos de los átomos:

$$\vec{F}_{\beta} = -\frac{\partial \mathcal{E}(\{\vec{R}_{\alpha}\})}{\partial \vec{R}_{\beta}} \quad (2.4)$$

Con esto los átomos se mueven a través de trayectorias clásicas pero la fuerza viene determinada por un cálculo de la estructura cuántica de los electrones del sistema.

Para poder calcular la estructura electrónica se pueden utilizar diversos métodos, algunos se basan en la función de onda como la aproximación de Hartree-Fock y otros métodos derivados de este. Estos métodos son muy pesados computacionalmente y hacen que sea muy difícil poder hacer dinámicas. También, como se verá en la siguiente sección, se puede resolver el sistema a partir de la densidad electrónica con el teorema de Hohenberg y Kohn que facilita mucho los cálculos.

### 2.2.2. Método QM: DFT

En 1964 Hohenberg y Kohn demostraron que un sistema de muchos electrones sometido a un potencial externo  $V_{ext}(\vec{r})$ , que en nuestro caso serán el potencial debido a los núcleos de los átomos, puede ser descrito por un funcional de la densidad electrónica  $E = E(\rho)$  y que para el caso del estado fundamental  $\rho_o(\vec{r})$  la energía es un mínimo. De forma que la energía del sistema se puede escribir de la siguiente forma

$$E^{HK}[\rho(\vec{r})] = T[\rho(\vec{r})] + E_{ee}[\rho_o(\vec{r})] + \int V_{ext}(\vec{r})\rho_o(\vec{r})d\vec{r} + E_{i-i} \quad (2.5)$$

donde  $T[\rho(\vec{r})]$  es la energía cinética de los electrones,  $E_{ee}[\rho_o(\vec{r})]$  es la energía electrón-electrón y  $E_{i-i}$  es la energía de interacción ion ion definida por:

$$E_{i-i} = \frac{e^2}{2} \sum_{i,j} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}. \quad (2.6)$$

Posteriormente Kohn y Sham propusieron un método para remplazar el problema de las partículas interactuantes por un problema auxiliar de partículas independientes donde efectos de muchos cuerpos están incluidos en el funcional de canje-correlación.<sup>11</sup> De forma que ahora la

energía queda descrita por la ecuación:

$$E^{KS}[\rho(\vec{r})] = T_0[\rho(\vec{r})] + E_H[\rho(\vec{r})] + \int \rho(\vec{r}) V_{ion} + E_{i-i} + E_{XC}[\rho(\vec{r})] \quad (2.7)$$

donde  $T_0[\rho(\vec{r})]$  es la energía cinética de los electrones del problema auxiliar de electrones independientes con densidad electrónica  $\rho(\vec{r})$ ,  $E_H$  es la energía de Hartree debido a la repulsión electrónica

$$E_H[\rho(\mathbf{r})] = \frac{e^2}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3r d^3r', \quad (2.8)$$

y  $E_{XC}$  es el término de energía del canje-correlación que se puede escribir como:

$$E_{XC} = T[\rho] - T_0[\rho] + E_{ee}[\rho] - E_H[\rho] \quad (2.9)$$

Y nos queda un problema auxiliar para resolver el canje correlación, donde:

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V_{KS}(\vec{r}) \right] \psi_i = \epsilon_i \psi_i \quad (2.10)$$

con la energía cinética  $T_0[\rho]$

$$T_0[\rho] = \sum_i \langle \psi_i | -\frac{\hbar^2}{2m} \nabla^2 | \psi_i \rangle f_i \quad (2.11)$$

donde  $f_i$  es la ocupación de los estados  $\psi_i$ . Y la densidad electrónica es:

$$\rho = \sum_i f_i |\psi_i(\vec{r})|^2 \quad (2.12)$$

Por lo tanto tenemos un potencial de Khom-Sham para la nube electrónica

$$V_{KS} = V_{ext} + V_H + V_{XC} \quad (2.13)$$



$V_{ext}$  es la potencial externo que en nuestro caso sera la interacción de los electrones con los núcleos de los átomos:

$$V_{ext}(\mathbf{r}) = \sum_{\alpha} \frac{-eZ_{\alpha}}{|\mathbf{r} - \mathbf{R}_{\alpha}|}, \quad (2.14)$$

( $eZ_{\alpha}$  es la carga del núcleo del átomo  $\alpha$  en la posición  $\mathbf{R}_{\alpha}$  y  $-e$  es la carga del electrón). Para simplificar los cálculos de densidad electrónica se puede utilizar pseudopotenciales. Estos pseudopotenciales a parte del núcleo del sistema contienen los electrones internos de forma que el cálculo de la densidad electrónica se realiza solo para los electrones de valencia que son los que intervienen en la mayoría de los procesos químicos.. De esta forma, se reduce el coste computacional del cálculo al tener que trabajar con una cantidad menor de electrones. El potencial de hartree sería:

$$V_H(\mathbf{r}) = \frac{\delta E_H}{\delta \rho(\mathbf{r})} = e \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r' \quad (2.15)$$

Y por último el canje-correlación  $V_{xc}$  es el funcional derivado de la energía del canje-correlación de DFT:  $E_{xc}[\rho(\mathbf{r})]$ ,<sup>11</sup>

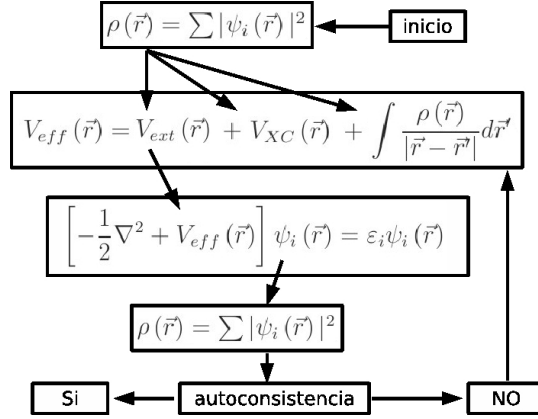
$$V_{xc} = \delta E_{xc} / \delta \rho(\mathbf{r}). \quad (2.16)$$

Y la ecuacion de Schrödinger DFT

$$\hat{H}\psi_i(\mathbf{r}) = \left( -\frac{\hbar^2}{2m_e} \nabla^2 + V_{ion}(\mathbf{r}) + V_{ee}(\mathbf{r}) + V_{xc}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}). \quad (2.17)$$

Estas ecuaciones se resuelven de forma auto-consistente donde se presupone una densidad electrónica  $\rho_{in}$  para el potencial  $V_{KS}$  y da como resultado una nueva densidad electrónica  $\rho_{out}$ , ver esquema en figura 2.1.

El mayor problema de DFT es que desconocemos la forma que tiene el término de canje-correlación  $E_{xc}$ . Para calcular este termino se pueden utilizar distintas aproximaciones para los funcionales de caje-correlación.



**Figura 2.1** Esquema de la autoconsistencia en DFT.

Las mas utilizadas son LDA 'Local-density approximation' donde su valor depende solo de la densidad electrónica en cada punto del espacio<sup>26</sup>. En la aproximación GGA 'Generalized Gradient Approximation' donde además depende de las derivadas de la densidad como BLYP<sup>27,28</sup> y PBE<sup>29</sup>. También hay funcionales híbridos como B3LYP<sup>30</sup> o PBE0<sup>31</sup> donde se resuelve de forma exacta parte del canje-correlación.

Una vez resueltas estas ecuaciones la energía total y las fuerzas se calculan a traves de:

$$E_{tot}[\rho(\mathbf{r})] = E_{sum} - E_H[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] - \int \rho(\mathbf{r}) V_{xc}[\rho(\mathbf{r})] d^3r + E_{i-i}; \quad (2.18)$$

$$\mathbf{F}_\alpha = -\frac{\partial E_{tot}}{\partial \mathbf{R}_\alpha}. \quad (2.19)$$

En la ecuación (2.18), la  $E_{sum}$  es una suma sobre los autovalores ocupados,  $E_{sum} = \sum_{i \in occ} \epsilon_i$ , de la autoconsistencia de la ecuación (2.17)

### 2.2.3. Método QM: FIREBALL

FIREBALL es un método DFT que está especialmente diseñado para el estudio de dinámicas en sistemas complejos. Sus características principales son: (i) el uso de orbitales numéricos de tipo atómico de corto alcance; (ii) está totalmente formulado en espacio real y no necesita el uso de cuadrículas para la densidad de carga; (iii) la autoconsistencia se realiza sobre la ocupación de los orbitales; (iv) hay un esquema de tabulación interpolación para que todas interacciones que se necesitan durante la simulación estén calculadas anteriormente y guardadas en tablas.

En el método FIREBALL se usa una variante del método de Khon-Sham, conocida como funcional de Harris-Foulkes.<sup>32,33</sup> En esta variante la energía para un sistema de electrones para una densidad de carga de entrada se escribe como:

$$E^{Harris}[\rho_{in}] = \sum_i \epsilon_i - E_{ee}[\rho_{in}] + E_{XC}[\rho_{in}] - \int \rho_{in}(\vec{r}) V_{XC}[\rho_{in}(\vec{r})] d^3r, \quad (2.20)$$

donde  $\epsilon_n$  son los autovalores ocupados del hamiltoniano efectivo de un electrón:

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V(\rho_{in}) \right] \psi_i = \epsilon_i \psi_i \quad (2.21)$$

con

$$V(\rho_{in}) = V_{ext}(\vec{r}) + \int \frac{\rho_{in}(\vec{r}') d^3r'}{|\vec{r} - \vec{r}'|} + V_{XC}(\rho_{in}(\vec{r})) \quad (2.22)$$

En estas ecuaciones la  $E^{Harris}$  está definida en función de  $\rho_{in}$  a diferencia de las de Khon-Sham que están en función de  $\rho$  que es la salida de la densidad electrónica. Esta es una característica importante del funcional de Harris-Foulkes ya que te permite utilizar  $\rho_{in}$  como una suma de densidades localizadas en las posiciones de los átomos. De esta forma la interacción coulombica electrón-electrón se queda en términos de tres centros lo que facilita mucho el cálculo.

FIREBALL usa una base de orbitales numéricos localizados<sup>20,25</sup>  $\phi_\mu$ , para resolver las ecuaciones de Hohn-Sham:

$$\psi_i(\mathbf{r}) = \sum_{\mu} c_{i\mu} \phi_{\mu}(\mathbf{r} - \mathbf{R}_{\alpha}); \quad (2.23)$$

donde  $\mu$  es un índice combinado,  $\mu \equiv (\alpha, l, m)$ , en el cual nos referimos a un átomo particular  $\alpha$  (en  $\mathbf{R}_{\alpha}$ ), con una subcapa  $l$  (e.g.  $3s, 4s, 3p, 3d$ ), y un momento angular  $m$  (e.g.  $p_x, p_y, p_z$ ).

En el método FIREBALL estas ecuaciones la densidad electrónica  $\rho(\mathbf{r})$ <sup>32,33</sup> esta determinada de forma autoconsistente en términos de los números de ocupación de los orbitales<sup>19</sup>. La densidad electrónica  $\rho(\mathbf{r})$  también se escribe en términos de los orbitales  $\phi_{\mu}$  como una suma de densidades de tipo atómicas,  $\rho_{\alpha}(\mathbf{r})$ , para los diferentes átomos del sistema:<sup>19</sup>

$$\rho^{in}(\mathbf{r}) = \sum_{\alpha} \rho_{\alpha}(\mathbf{r}) = \sum_{\mu} n_{\mu} |\phi_{\mu}(\mathbf{r} - \mathbf{R}_{\alpha})|^2. \quad (2.24)$$

De forma práctica las densidades atómicas se aproximan con una simetría esférica en torno a la posición de cada átomo  $\mathbf{R}_{\alpha}$  es decir,  $n_{\alpha lm} = n_{\alpha l m'}$ , para los diferentes valores de  $m$  para cada capa  $\alpha$  l. Por ejemplo la densidad electrónica  $\rho(\mathbf{r})$  en la ecuación (2.24) se determina de forma autoconsistente en términos de los números de ocupación de los orbitales  $n_{\mu}$ .<sup>19,34,35</sup> Para ello, se obtienen números de ocupación  $n_{\mu}^{out}$  proyectando la densidad electrónica  $\rho^{out}(\vec{r})$

$$\rho^{out}(\mathbf{r}) = \sum_{i \in occ} |\psi_i(\mathbf{r})|^2 = \sum_i f_i |\psi_i(\mathbf{r})|^2 \quad (2.25)$$

en la forma dada en la ecuación 2.24,  $\rho^{in}$

$$\rho^{out}(\mathbf{r}) \longrightarrow \sum_{\mu} n_{\mu}^{out} |\phi_{\mu}(\mathbf{r} - \mathbf{R}_{\alpha})|^2. \quad (2.26)$$

La buena eficiencia computacional de FIREBALL depende tanto del uso de una base optimizada de orbitales numéricos de tipo atómico como de la representación de  $\rho(\mathbf{r})$  como una suma de densidades de tipo atómico (ecuación 2.24).<sup>36</sup> Ya que de esta forma no son necesarias las integrales de cuatro centros para resolver la ecuación (2.17) o para el cálculo de la  $E_{tot}$  y de las fuerzas  $\mathbf{F}_\alpha$ . Esto también permite tabular todas las integrales de tres centros necesarias y guardarlas en tablas de datos de dos dimensiones.<sup>16,20</sup> Se utiliza un esquema tabulación-interpolación donde todos los elementos de matriz necesarios para hacer una simulación MD son guardados en archivos. Los archivos se pueden leer al principio de la simulación y se guardan en la memoria del ordenador. Este esquema de tabulación-interpolación también se utiliza para los términos de canje-correlación, que son altamente no lineales, con las diferentes aproximaciones que se han desarrollado.<sup>17,18,20,37</sup>

Se pueden utilizar diferentes métodos para proyectar [flecha en la ecuación (2.26)] la densidad electrónica en forma de las ecuaciones (2.24). Una buena forma de hacer esta proyección es a través de orbitales Löwdin<sup>19,38,39</sup>:

$$\varphi_\mu(\vec{r}) = \sum_{\nu} (S^{-1/2})_{\mu\nu} \phi_\nu(\vec{r} - \vec{R}_\alpha) \quad (2.27)$$

donde S es la matriz de solape,  $S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle = \int \phi_\mu^*(\vec{r} - \vec{R}_\alpha) \phi_\nu(\vec{r} - \vec{R}_\alpha) d^3\vec{r}$ . Los orbitales Löwdin forman una base ortogonal,  $\langle \varphi_\mu | \varphi_\nu \rangle = \delta_{\mu\nu}$ . Los orbitales  $\varphi_\mu$  son los orbitales ortonormales mas parecidos a los orbitales originales (no ortogonales),  $\phi_\mu$ , en el sentido que minimizan el error cuadrático medio.

$$\sum_{\mu} \int |\varphi_\mu - \phi_\mu|^2 d^3\vec{r} \quad (2.28)$$

Las ecuaciones (2.17-2.26) se resuelven de forma iterativa de manera autoconsistente tal que  $n_\mu^{out}$  y  $n_\mu^{in}$  deben coincidir.<sup>19</sup> En cada una de estas iteraciones los autovectores  $\psi_i$  se obtienen de la diagonalización de la

matriz del hamiltoniano  $N \times N$  en la ecuación (2.17),  $H_{\mu,\nu} = \langle \phi_\mu | \hat{H} | \phi_\nu \rangle$ , donde  $N$  es el número total de orbitales.

Para el cálculo de las fuerzas  $\mathbf{F}_\alpha$ , ecuación (2.19), el término mas costoso de calcular es contribución de  $E_{sum}$ . Para ello, se usa una variación de el teorema Hellmann-Feynman<sup>20</sup> que nos permite calcular este término en función de las derivadas (gradientes) de los elementos de matriz del hamiltoniano.

Como se ha mencionado, estos elementos de matriz se calculan usando tablas de datos pre-calculadas y el esquema de tabulación-interpolación, de tal manera que en función de las derivadas de los elementos de matriz  $H_{\mu\nu}$ ,  $\frac{\partial H_{\mu\nu}}{\partial \vec{R}_\alpha}$ , con respecto a las posiciones de los átomos,  $\vec{R}_\alpha$

$$\begin{aligned} \frac{\partial}{\partial \vec{R}_\alpha} \left( \sum_i f_i \epsilon_i \right) &= \sum_i \frac{\partial}{\partial \vec{R}_\alpha} \langle \phi_i | \hat{h} | \phi_i \rangle = \\ &= \sum_{i\mu\nu} c_{i\mu}^* c_{i\nu} \frac{\partial h_{\mu\nu}}{\partial \vec{R}_\alpha} - \sum_{i\mu\nu} e_i c_{i\mu}^* c_{i\nu} \frac{\partial S_{\mu\nu}}{\partial \vec{R}_\alpha} \end{aligned} \quad (2.29)$$

también se pueden calcular antes y guardarlos en tablas de datos de forma que los términos derivativos que se necesitan  $\partial H_{\mu,\nu} / \partial \mathbf{R}_\alpha$  esten accesibles, ver ref<sup>20</sup> para los detalles.

#### 2.2.4. FIREBALL / AMBER

En los métodos QM/MM se divide el sistema en una región químicamente activa (residuos del sitio catalítico y molécula reactante) y el ambiente (resto de la proteína, moléculas de agua, iones, etc). Normalmente la región químicamente activa suele tener alrededor de  $10^2$  átomos y se utiliza QM para el cálculo mientras que el ambiente tiene unos  $\sim 10^5 - 10^6$  átomos lo cual obliga a calcularlo con MM.

Al estudiar reacciones en biomoléculas es particularmente importante tener en cuenta el efecto electrostático del ambiente sobre la estructura electrónica de la región activa.

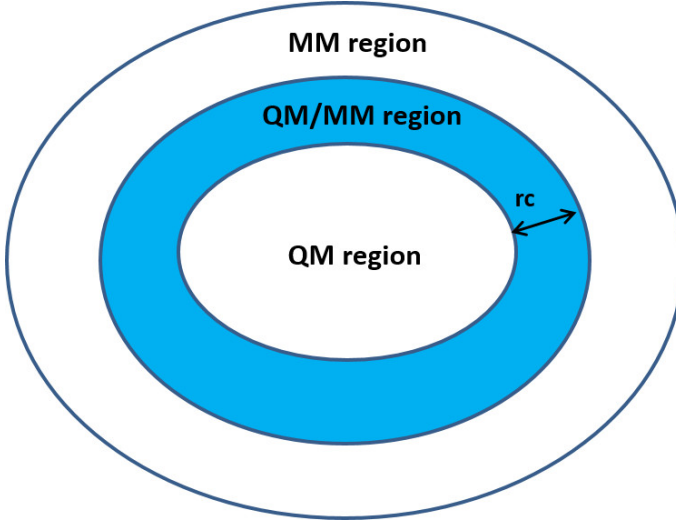
En el método QM/MM FIREBALL / AMBER los cálculos se tiene que separar en distintas regiones donde cada una de ellas tendrá una aproximación distinta, en concreto se separaran en tres regiones distintas (ver fig. 2.2):

- Región QM, átomos en la región químicamente activa. Esta región se describe con el método QM FIREBALL. Se verán afectados por la interacción electrostática con los átomos de MM.
- Región QM/MM, estos átomos están descrito por el campo de fuerza MM al que se le añade un potencial debido a la interacción con los átomos de QM.
- Región MM, átomos de MM que están mas lejos de la región QM y que solo son simulados mediante el campo de fuerzas MM.

En cada una de estas regiones tiene que calcularse las fuerzas y energías de forma diferente. El programa al empezar cada paso tiene que definir qué átomos están en cada región y separarlo en distintas matrices. En el caso de la región QM los átomos involucrados van a ser los mismos, en principio, durante toda la dinámica por lo que esa matriz sera definida al principio del programa. En cambio los átomos que están en la región QM/MM y MM cambian en cada paso ya que se recalcula si están dentro de la distancia  $r_c$  de cualquier átomo de QM.

En esta implementación FIREBALL/AMBER la interacción electrostática entre la región QM y QM/MM introduce una nueva contribución  $\delta H_{\mu,\nu}$  a los elementos de matriz del Hamiltoniano de la región QM ec 2.17.

$$H_{\mu,\nu} = H_{\mu,\nu}^A + \delta H_{\mu,\nu}^E. \quad (2.30)$$



**Figura 2.2** Diferentes regiones en una simulación de QM/M

$H_{\mu,\nu}^A$  es la contribución de los átomos en la región activa a los elementos de matriz del hamiltoniano.  $\delta H_{\mu,\nu}^E$  es la contribución de los elementos de matriz  $\delta H_{\mu,\nu}$  debido al potencial electrostático de las cargas parciales  $eq_K$  de los átomos en el entorno (región QM/MM).

$$\delta H_{\mu,\nu}^E = \sum_k h_{\mu,\nu}^k = \sum_k -eq_k \int \frac{\phi_\mu(\mathbf{r} - \mathbf{R}_\alpha) \phi_\nu(\mathbf{r} - \mathbf{R}_\beta)}{|\mathbf{r} - \mathbf{R}_k|} d^3\mathbf{r}. \quad (2.31)$$

$\phi_\mu(\mathbf{r} - \mathbf{R}_\alpha)$  y  $\phi_\nu(\mathbf{r} - \mathbf{R}_\beta)$  son los orbitales de la base de tipo atómico que están centrados en el átomo que está en la posición  $\mathbf{R}_\alpha$  y  $\mathbf{R}_\beta$ .  $h_{\mu,\nu}^k$  es la contribución a los elementos de matriz del hamiltoniano  $H_{\mu,\nu}$  debido al potencial electrostático de una carga parcial  $eq_k$  en  $\mathbf{R}_k$ , ver figura 2.3. Si expandimos  $1/|\mathbf{r} - \mathbf{R}_k|$  en polinomios de Legendre tenemos

$$h_{\mu,\nu}^k = \frac{S_{\mu\nu}}{|\vec{R}_k - \vec{R}_0|} + \frac{1}{|\vec{R}_k - \vec{R}_0|^3} (\vec{R}_k - \vec{R}_0) \cdot \vec{p}_{\mu\nu}(\vec{R}_0) + \dots \quad (2.32)$$



donde  $\rho_{\alpha\beta}$  es el momento dipolar;  $\vec{p}_{\alpha\beta}(\vec{R}_0) = 0 \rightarrow \vec{R}_0$

$$\vec{p}_{\mu\nu}(\vec{R}_0) = \int \phi_{\mu}^*(\vec{r} - \vec{R}_{\mu}) \phi_{\nu}(\vec{r} - \vec{R}_{\nu}); \quad (2.33)$$

tomamos el origen en  $\vec{R}_m = \frac{\vec{R}_{\alpha} + \vec{R}_{\beta}}{2}$

$$\vec{p}_{\mu\nu}(\vec{R}_0) = \vec{p}_{\mu\nu}(\vec{R}_m) + (\vec{R}_m - \vec{R}_0) S_{\mu\nu} = 0 \rightarrow \vec{R}_0 = \vec{R}_m + \frac{\vec{p}_{\mu\nu}(\vec{R}_m)}{S_{\mu\nu}} \quad (2.34)$$

Y obtenemos que

$$h_{\mu,\nu}^k = \frac{S_{\mu\nu}}{\vec{R}_k - \vec{R}_m - \frac{\vec{p}_{\mu\nu}(\vec{R}_m)}{S_{\mu\nu}}} \quad (2.35)$$

$S_{\mu,\nu}$  es el solape:

$$S_{\mu,\nu} = \int \phi_{\mu}(\mathbf{r} - \mathbf{R}_{\alpha}) \phi_{\nu}(\mathbf{r} - \mathbf{R}_{\beta}) d^3\mathbf{r}, \quad (2.36)$$

y  $p_{\mu,\nu}$  es la componente a lo largo de la dirección  $(\mathbf{R}_{\beta} - \mathbf{R}_{\alpha})$  del momento dipolar  $\mathbf{P}_{\mu,\nu}$  con respecto al punto medio,  $\mathbf{R}_m$ , entre ambos átomos  $\alpha$  y  $\beta$ :

$$\mathbf{P}_{\mu,\nu} = \int [\mathbf{r} - \mathbf{R}_m] \phi_{\mu}(\mathbf{r} - \mathbf{R}_{\alpha}) \phi_{\nu}(\mathbf{r} - \mathbf{R}_{\beta}) d^3\mathbf{r}, \quad (2.37)$$

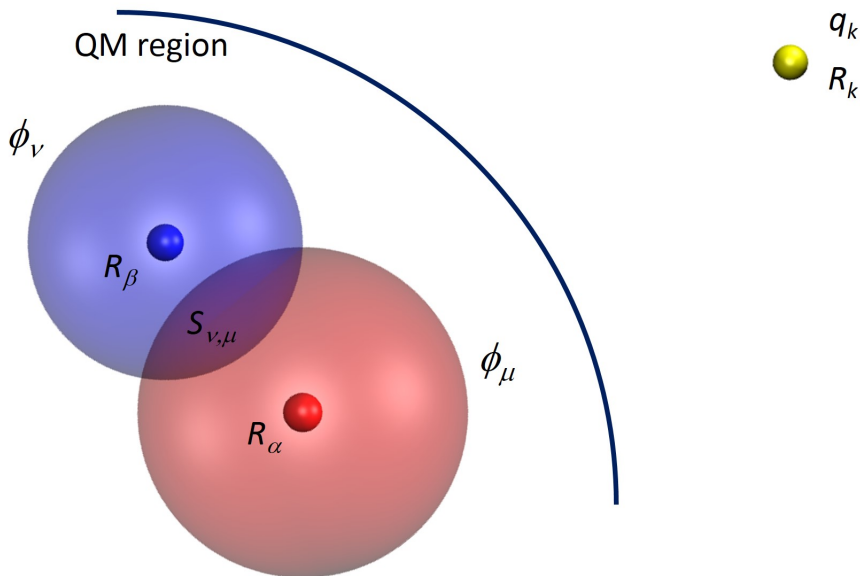
$(\mathbf{R}_m = (\mathbf{R}_{\alpha} + \mathbf{R}_{\beta})/2)$ . En la aproximación de la ecuación (2.38) se incluye la física de un monopolo y dipolo a larga distancia.<sup>19</sup>

En el mismo orden de aproximación, se puede escribir esta ecuación de la siguiente forma

$$h_{\mu,\nu}^k = -eq_k \left( \frac{g_{\mu}}{|\mathbf{R}_k - \mathbf{R}_{\alpha}|} + \frac{g_{\nu}}{|\mathbf{R}_k - \mathbf{R}_{\beta}|} \right), \quad (2.38)$$

con

$$g_{\mu} = \frac{1}{2} S_{\mu,\nu} - \frac{p_{\mu,\nu}}{|\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}|}, \quad (2.39)$$



**Figura 2.3** Esquema representativo de la interacción electrostática entre una carga de MM,  $eq_k$ , en  $\mathbf{R}_k$  y el solape de carga,  $-eS_{\mu,v}$ , que produce la contribución del elemento de matriz  $h_{\mu,v}^k$  al hamiltoniano de QM  $H_{\mu,v}$  entre los orbitales  $\phi_\mu(\mathbf{r} - \mathbf{R}_\alpha)$  y  $\phi_v(\mathbf{r} - \mathbf{R}_\beta)$ , ver ecuaciones (2.31) and (2.38).

$$g_v = \frac{1}{2}S_{\mu,v} + \frac{p_{\mu,v}}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}; \quad (2.40)$$

El potencial electrostático debido al entorno QM/MM también interactúa con la carga del núcleo (o pseudopotencial) de los átomos de QM,  $eZ_\alpha$ , dando lugar a la siguiente contribución de energía:<sup>15</sup>

$$E_{QM-MM}^{nuc} = \sum_{\alpha,k} \frac{e^2 q_k Z_\alpha}{|\mathbf{R}_k - \mathbf{R}_\alpha|}. \quad (2.41)$$

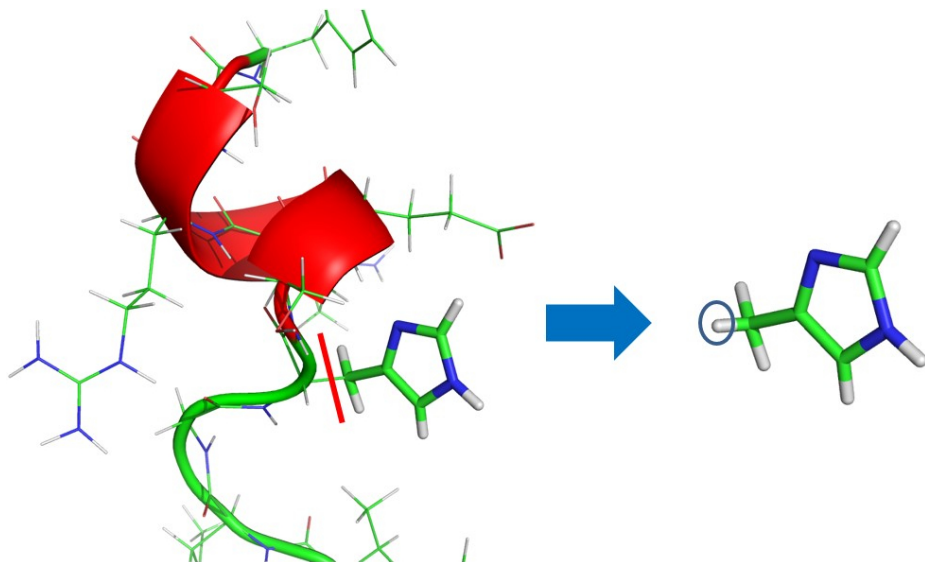
Con sus correspondientes fuerzas en  $\vec{R}_k$  y  $\vec{R}_\alpha$

Además del término de interacción electrostática en la interacción entre átomos en las regiones de QM y QM/MM hay otras contribuciones que se incluyen mediante las correspondientes contribuciones del campo de fuerzas de MM (terminos enlazantes y no-enlazantes).<sup>8</sup> En la mayoría de aplicaciones hay átomos de QM y QM/MM que están unidos por un enlace covalente. Estos enlaces hay que tratarlos de forma específica de forma que el átomo de QM este adecuadamente saturado en el cálculo de DFT. También hay términos no-enlazantes entre átomos de las regiones de QM y QM/MM que están descritos mediante los correspondientes parámetros del campo de fuerzas clásic. La interacción electrostática a más larga distancia para todos los átomos del sistema que se calcula usando un método *Particle Mesh Ewald*.<sup>40</sup>

En esta implementación hemos usado la aproximación de 'link atom' donde se añaden unos átomos al calculo de QM para saturar los enlaces rotos en la frontera entre átomos de QM y QM/MM. El átomo *link* se coloca a lo largo del vector que une a los correspondientes átomos de las regiones QM y QM/MM, ver figura 2.4. Para saturar el átomo link se define una distancia  $d_{L-QM}$  que es una variable que añade el usuario y suele ser la distancia típica de enlace entre el átomo añadido, típicamente hidrógeno, y el de QM. La coordenada del átomo link es:

$$\vec{r}_L = \vec{r}_{QM} + d_{L-QM} \frac{\vec{r}_{MM} - \vec{r}_{QM}}{|\vec{r}_{MM} - \vec{r}_{QM}|} \quad (2.42)$$

Donde  $\vec{r}_L$ ,  $\vec{r}_{QM}$  y  $\vec{r}_{MM}$  son los vectores de posición de los átomos de QM y MM respectivamente.



**Figura 2.4** Esquema representativo del corte entre la región de QM y MM cuando hay un enlace covalente y la colocación del átomo link.

Las fuerzas que se obtiene para el átomo *link* en el cálculo de QM, se añade a los dos átomos, QM y MM, entre los cuales había roto en enlace de la siguiente forma:

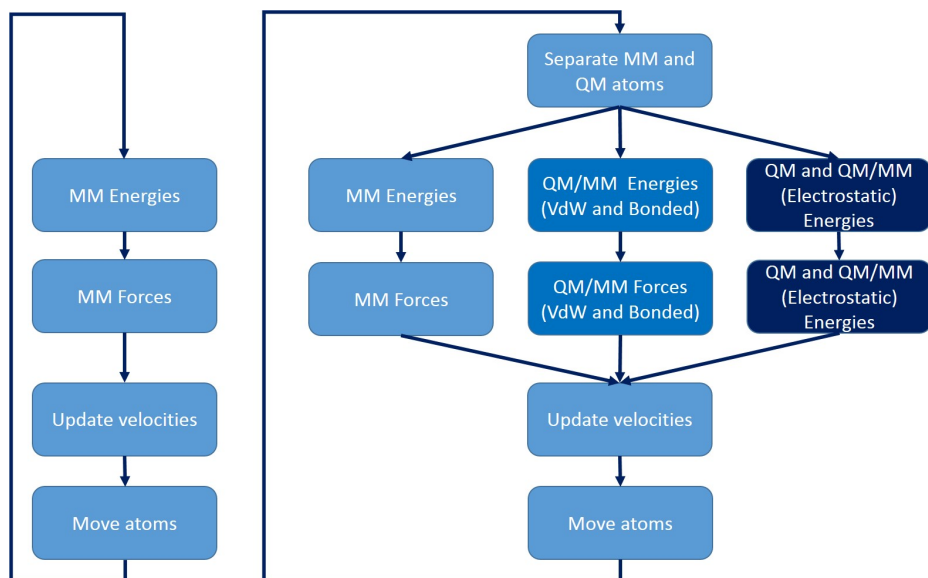
$$\begin{aligned}\vec{F}_{QM} &= -\frac{\partial E(\vec{r}_L)}{\partial x_{QM}} = -\nabla_L E \cdot \frac{\partial \vec{r}_L}{\partial r_{QM}} = \vec{F}_L \frac{\partial \vec{r}_L}{\partial r_{QM}} \\ \vec{F}_{MM} &= \vec{F}_L \frac{\partial \vec{r}_L}{\partial r_{MM}}\end{aligned}\tag{2.43}$$

Donde  $\vec{F}_{QM}$  y  $\vec{F}_{MM}$  son las fuerzas añadidas a los átomos de QM y MM respectivamente y  $\vec{F}_L$  es la fuerza sobre el átomo link en el cálculo de QM. Y sus derivadas parciales se pueden expresar como:

$$\begin{aligned}\frac{\partial \vec{r}_L}{\partial x_{QM}} &= \left(1 - \frac{d_{L-QM}}{d_{MM-QM}}\right) \vec{i} + \frac{d_{L-QM}(X_{MM} - x_{QM})}{d_{MM-QM}^3} (\vec{r}_{MM} - \vec{r}_{QM}) \\ \frac{\partial \vec{r}_L}{\partial x_{MM}} &= \frac{d_{L-QM}}{d_{MM-QM}} \vec{i} - \frac{d_{L-QM}(X_{MM} - x_{QM})}{d_{MM-QM}^3} (\vec{r}_{MM} - \vec{r}_{QM})\end{aligned}\quad (2.44)$$

En la implementación de FIREBALL/AMBER se utilizó como base el código de Walker et al.<sup>40</sup> para el programa *sander*. Nos podemos hacer una idea de como funciona una dinámica clásica con un diagrama de flujo lo mas sencillo posible. En el caso de una dinámica clásica en la que no hay que dividir el sistema se calcula de forma secuencial los distintos pasos (ver 2.5), primero se calcula la energía después la fuerza y por último se cambian las velocidades y se mueven los átomos.

Para una dinámica QM/MM la simulación esta dividida en varias regiones cada una con una aproximación distinta. De forma que el diagrama de flujo cambia, lo primero que hay que hacer es la división de las regiones y después de forma independiente en cada región hay que calcular las energías y fuerzas en cada átomo. Con los átomos que están descritos de forma clásica se procederá de la misma forma que hemos descrito anteriormente para una dinámica clásica. Por otro lado se calcula la energía de los átomos de QM de forma autoconsistente. Para calcular esta energía hay que tener en cuenta que los átomos de QM necesitan saber donde están todos los átomos en la región QM/MM ya que la energía dependerá de la posición y la carga parcial de estos átomos. Después, al calcular la fuerza en los átomos de la región QM hay que tener en cuenta que los átomos de la región QM/MM estaban modificando la estructura electrónica. Esta interacción, con sus correspondientes fuerzas, se calcula en la misma sección que los átomos de la región QM ya que afecta directamente al



**Figura 2.5** Diagrams de flujo simplificados para una dinámica clásica a la izquierda y para una dinámica QM/MM a la derecha.

hamiltoniano. Para los átomos de la región QM/MM a parte de la fuerza por la interacción electrostática que ya hemos calculado antes hay que calcular la energía y la fuerza para las interacciones no-enlazantes (Lennard-Jones) y los términos de enlace que hemos cortado. (ver 2.5)

## 2.3. Ejemplo: efecto de las cargas de MM en la región de QM en un puente salino

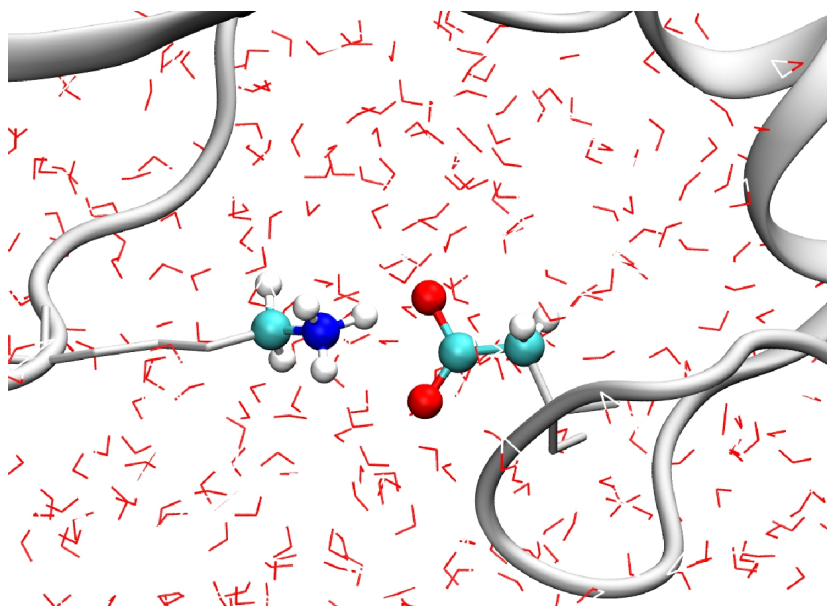
Como ejemplo de la importancia del potencial electrostático del entorno en un centro activo vamos a ver como el tener en cuenta o no esta interacción puede modificar el resultado de una dinámica.

La interacción entre dos residuos cargados suele tener un papel muy importante en la estructura y función de las proteínas. Esta interacción en vacío suele ser muy grande pero cuando se encuentran en un

ambiente polar (aguas o residuos polares) esta energía de interacción se puede atenuar bastante. Para probar cómo afecta el efecto de las cargas parciales de MM en nuestra región de QM en la implementación vamos a simular un puente salino formado entre un Aspártico y una Lisina. Esta simulación se hará con 'Electrostatic embedding' (teniendo en cuenta como las cargas afectan al hamiltoniano de QM, ver ecuacion 2.32 ) y con 'Mechanical embedding' (donde las cargas parciales no afectan a la estructura electrónica pero si a la energía total y las fuerzas). El puente salino que se estudia es uno formado entre el Asp 69 y la Lys 219 de la transcriptasa inversa del virus de sida HIV-1 que se sabe que tiene un efecto importante en la hypersusceptibilidad del AZT.<sup>41</sup>

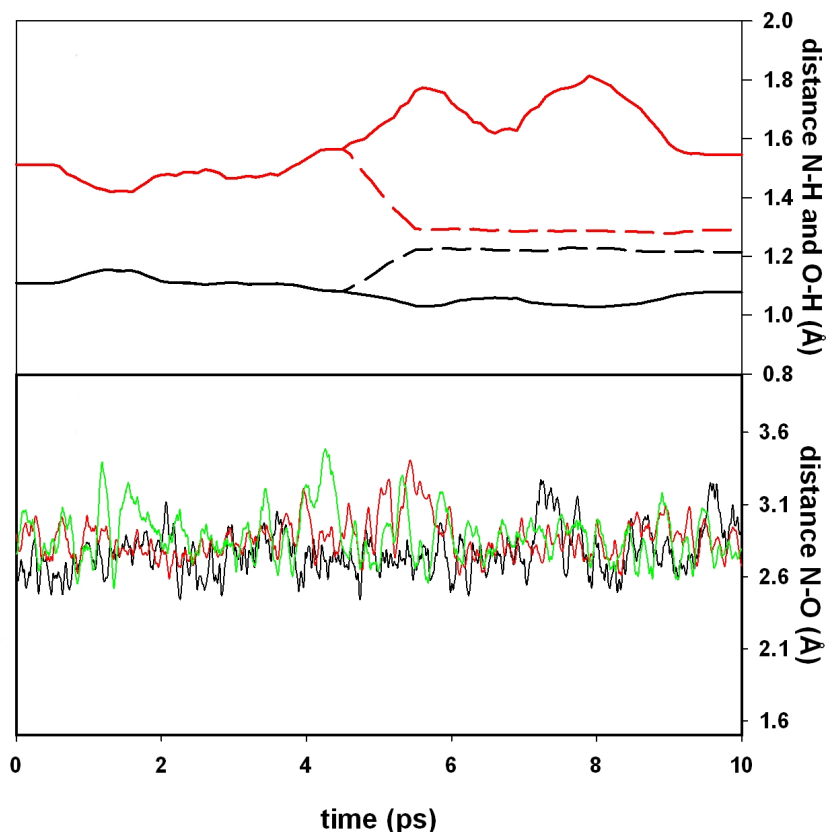
Empezando desde una región estable de una simulación clásica de MD para esta proteína se hizo una dinámica QM/MM de 20 000 pasos ( $\Delta t = 0.5$  fs), donde se definió la región de QM con los átomos implicados en este puente salino (ver figura 2.6). En las simulaciones con *Electrostatic embedding* obtuvimos que las distancias entre el H y el O del puente salino estaban entre 1,6 – 1,7 Å que son las esperadas para este tipo de interacción (ver la linea solida roja en la figura 2.7 a). En cambio si desactivábamos la interacción, a partir de los 5 ps, de las cargas parciales sobre nuestro Hamiltoniano de QM (lineas discontinuas en la figura 2.7 a) se ve cómo la distancia entre el H y los oxígenos del Asp se reduce y pasa a tener un comportamiento oscilatorio como consecuencia de una interacción mas fuerte como podríamos esperar que fuera en el vacío.

También analizamos la primera capa de solvatación de este puente salino y vimos que de media había cuatro aguas alrededor del grupo amonio que estabilizaban la interacción, para el grupo carboxílico el oxígeno involucrado en el puente de hidrógeno tenía dos aguas y el otro oxígeno en la primera capa tenía tres aguas. Se vio que de media había la misma cantidad de aguas en la primera capa tanto en simulaciones de dinámica molecular clásica como en simulaciones QM/MM.



**Figura 2.6** Simulación de dinámica Molecular entre para la interacción entre Aspártico y Lisina en un puente salino. Sistema simulado: La región de QM esta representada en *ball and sticks*, el cadena principal de la proteína en *ribbons* y las aguas en líneas rojas.





**Figura 2.7** Simulación de dinámica Molecular entre para la interacción entre Aspártico y Lisina en un puente salino: (a) Distancias suavizadas a lo largo de la simulación de QM/MM entre el N y los átomos de H (líneas negras) y entre el O y los átomos de H (líneas rojas) que forman el puente salino: *electrostatic embedding* (líneas solidas) y *mechanical embedding* (líneas discontinuas). (b) Distancia entre el O y el N que forman el puente salino para las tres simulaciones: (línea negra) con el Aspártico y la Lisina en la región de QM; (línea roja) el aspártico en la región de QM región la lisina en la región de MM; (línea verde) lisina en la región de QM y aspártico en la región de MM.

Para continuar haciendo pruebas del efecto de la interacción electrostática en el hamiltoniano de QM hicimos pruebas donde solo estaba uno de los residuos del puente salino en QM y el otro permanecía en MM. Las pruebas se hicieron tanto con el aspártico en QM y la lisina en MM como viceversa. En la figura 2.7 b se ve como la distancia entre los átomos de N y O del puente salino permanecían estables y variando en torno al mismo valor 2.6-3.1 Å, tanto si solo uno de los residuos estaba en cuántica como si estaban los dos.

Este resultado indica que la interacción electrostática incluida en la implementación FIREBALL/AMBER reproduce de forma correcta el efecto que un puente salino tiene al estar rodeado de un ambiente polar con el agua y los residuos de la proteína.

## CAPÍTULO 3

### OPTIMIZACIÓN DE ORBITALES FIREBALL PARA SISTEMAS BIOLÓGICOS

---

En el capítulo anterior desarrollamos un método QM/MM para simular biomoléculas con el código DFT FIREBALL el cual necesita una base de orbitales de tipo atómico. Además se explicó que la eficiencia computacional del código dependía de la complejidad de esta base. En este capítulo se desarrollará una base mínima para poder reducir el coste computacional y se intentará mantener la mayor precisión posible. Los orbitales de tipo atómico de FIREBALL están descritos con ciertos parámetros, que se desarrollarán posteriormente, que nos van a permitir modificar la base y buscar el mejor balance posible entre la eficiencia computacional y la precisión. Ya que esta base está pensada para el estudio de biomoléculas para optimizar la base se utilizarán 'benchmarks' desarrollados para sistemas biológicos.

Primero explicaremos como funcionan los orbitales numéricos de FIREBALL y el significado de los distintos parámetros que se pueden modificar. Posteriormente se optimizarán los orbitales comparando nuestros resultados de energía y distancia como los 'benchmarks' desarrollados para estructuras biomoleculares.

Las biomoléculas son sistemas muy complejos que suelen tener un gran número de átomos, además los procesos en los que participan suelen ser complejos, reacciones enzimáticas, reacciones foto-químicas, etc. Por ello este balance entre eficiencia computacional y precisión es fundamental para poder estudiar estos sistemas de forma dinámica y tener suficiente precisión para poder describir de forma correcta el sistema.

## 3.1. Method

### 3.1.1. Orbitales numéricos tipo atómico

Normalmente los métodos de estructura electrónica con orbitales locales están definidos por una base del tipo:

$$\varphi_{\alpha n l m}(\mathbf{r}) = \phi_{\alpha n l}(r) Y_{l m}(\Omega) = \phi_{\alpha \mu}(r) Y_{l m}(\Omega) \quad (3.1)$$

( $\mu \equiv (n, l)$ ) con  $r = |\vec{r} - \vec{R}_{\alpha}|$ ;  $\alpha$  es el índice del átomo,  $\vec{R}_{\alpha}$  es la posición de átomo  $\alpha$  y  $Y_{l m}$  es el armónico esférico para los momentos angulares  $l$ ,  $m$  y el número atómico principal es  $n$ . Para construir los orbitales  $\phi_{\alpha \mu}$  se pueden utilizar diferentes estrategias. Las más comunes son orbitales de tipo Slater (STO) (cita), orbitales del tipo a Gaussianas (GTOs) (cita) y orbitales numéricos de tipo atómico (NAOS).

Los orbitales atómicos en sólidos y moléculas: No son lo mejor debido al incremento de energía cinética por el solape de las densidades atómicas. Optimización de los GTOs o STOs: cambian los exponentes. La optimización no es demasiado eficiente a causa de los potenciales confinantes

En cambio la optimización de los orbitales atómicos numéricos (NAOs) se puede obtener como una combinación lineal de orbitales de una base primitiva con muy buenos resultados. (Ver trabajo de Basanta) En un trabajo anterior Basanta se optimizó una base de Carbono e Hidrógeno

utilizando el 'mixing factor' y se creo una base simple que reproducía perfectamente

Los orbitales *fireball* de Sankey y Niklewski se obtienen al resolver el problema atómico con una condición de límite en la cual se obliga al orbital a ir a cero a partir de una distancia dada,  $R_c$ .

Estas bases se pueden mejorar de dos formas:

- (1) Añadiendo orbitales por átomo;
- (2) Optimizando los orbitales locales e intentando mantener reducido el número de orbitales.

Al añadir orbitales se mejora bastante la base a costa de aumentar significativamente el coste computacional y también hay que optimizar parte ya que sigue habiendo parámetros como el Radio de corte En la optimización de los orbitales locales se optimizan las funciones de onda radial  $\phi_{\alpha\mu}$  para cada capa electrónica de los átomos y se pueden obtener muy buenos resultados con un coste computacional bajo.

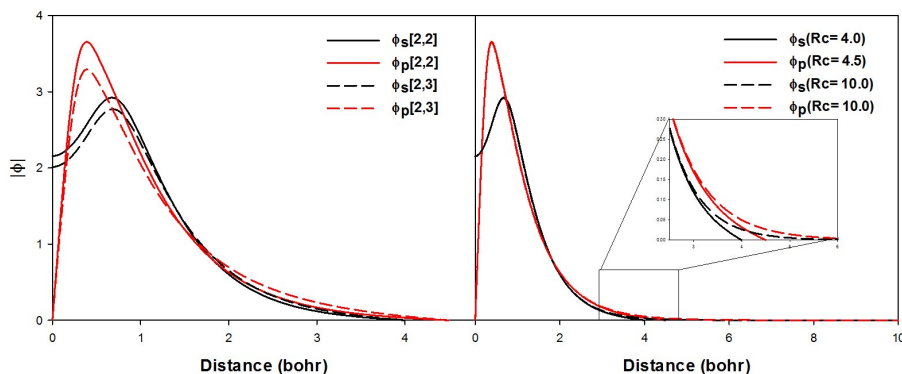
Hay varias formas de optimizar la función de onda radial de los átomos. Por ejemplo se puede cambiar la forma de la parte radial utilizando un potencial confinante del tipo:

$$V(r) = \begin{cases} V_0 \frac{\exp\left(-\frac{r_c - r_0}{r - r_0}\right)}{r_c - r} & r > r_0 \\ 0 & \end{cases} \quad (3.2)$$

para hacer que la cola de este orbital decaiga a cero de forma mas suave.

También se puede optimizar las funciones de onda de una base simple con una combinación lineal entre dos orbitales obtenidos con diferentes cálculos atómicos<sup>?</sup>.

$$\phi_{\alpha\nu}^{OPT}(r) = A[c_{\alpha\nu}\phi_{\alpha\nu}^0(r) + (1 - c_{\alpha\nu})\chi_{\alpha\nu}(r)]; \quad (3.3)$$



**Figura 3.1** Función de onda radial de las NAOs  $\phi_s$  (línea negra) y  $\phi_p$  (línea roja) para el átomo de Nitrógeno. (a) Gráfica de los orbitales en función de la distancia para diferentes cargas atómicas  $\{q_s, q_p\}$ . Orbitales para el átomo neutro ([2,3]; línea discontinua) y átomo optimizado ([2,2]; línea sólida), con  $R_c(s) = 4.0$  a.u. y  $R_c(p) = 4.5$  a.u. (b) Función radial con diferentes radios de corte. (i)  $R_c(s) = 4.0$  a.u.,  $R_c(p) = 4.5$  a.u. (línea sólida) y (ii)  $R_c(s) = R_c(p) = 10.0$  a.u. (línea discontinua); con cargas [2,2] para los orbitales s y p.

donde,  $\phi_{\alpha v}^0$  es el NAO obtenido del cálculo de un átomo neutro, y  $\chi_{\alpha v}$  de un cálculo de un ion 2+.  $c_{\alpha v}$  es una constante de mezcla y A es la constante de normalización.

En este capítulo se obtendrá una base practica de NAOs optimizada para el cálculo de sistemas biomoleculares. Para optimizar esta base se generaran diferentes funciones de onda radiales  $\phi_{\alpha v}$  para cada átomo  $\alpha$  variando la carga atómica  $\{q_\mu\}$  y el radio de corte  $\{R_c(\mu)\}$  al resolver las funciones de onda que generan los NAOs. Estos orbitales  $\phi_{\alpha v}[\{q_{\alpha\mu}\}]$  son generados con un cálculo atómico donde tienen un radio de corte dado  $\{R_c(\mu)\}$  y una carga  $\{q_{\alpha\mu}\}$  en cada capa atómica. Por ejemplo, para el C el orbital  $\phi_s[2,2]$  es el orbital s obtenido de un calculo atómico con 2 electrones en la capa s y 2 electrones en la capa p y unos radios de corte  $R_c(s)$  y  $R_c(p)$  en cada capa. Este procedimiento tiene la ventaja de permitir una gran flexibilidad para crear orbitales con funciones radiales

distintas. En la figura 3.1 se ve la función radial  $\phi_s, \phi_p$  para el átomo de N con diferentes cargas  $\{q_s, q_p\}$  y radios de corte  $R_c$ .

El átomo de H es un caso especial y sus funciones de onda se obtuvieron usando la ecuación 3.3 con  $\phi_s^0$  del átomo neutro de H y  $\chi_s$  de un ion  $\text{Li}^{+2}$ .

### 3.1.2. Optimización: Bases de datos y benchmarks

Para evaluar la calidad de la base se hicieron cálculos en distintos sets de datos para las biomoléculas mas relevantes donde se observaron distancias y energías. Para estos cálculos se utilizo el código FIREBALL con un funcional BLYP y la aproximación de McWeda. En todos los casos determinamos la desviación absoluta media (MAD) para cada set de datos comparando nuestro resultados con el 'benchmark'.

Los set utilizados fueron:

- **S66** Base de datos S66<sup>?</sup>. Una base para interacciones no covalentes con un buen balance entre las estructuras típicas que hay en biología. Hay 66 estructuras diferentes en las que se ha buscado la mejor geometría posible y se ha calculado la energía de interacción con métodos de química cuántica de alta precisión
- **IHB** Base de datos para puentes de hidrógeno iónicos<sup>?</sup>. También tiene geometrías y energías de interacción con métodos de química cuántica. En este caso las interacciones son puentes de hidrógeno mucho mas fuertes donde ha habido separación de cargas entre las moléculas.
- **COV** Base de datos de distancias de enlace. Las distancias están obtenidas de resultados experimentales, ver apéndice A. Se han buscado distancias típicas que existan en biomoléculas y se ha balanceado para que tengan un peso correcto todas las interacciones.

Con estas bases de datos calculamos la desviación absoluta media (MAD) para las energías (con optimización geométrica o con la del 'benchmark') y distancias para cada una de las bases y podemos hacer las siguientes mediciones:

- **S66 fix**: Energía de interacción en la base S66<sup>?</sup> entre las moléculas sin optimización geométrica.
- **S66 geo**: Energía de interacción en la base S66<sup>?</sup> entre las moléculas con optimización geométrica.
- **IHB fix**: Energía de interacción en la base IHB<sup>?</sup> entre las moléculas sin optimización geométrica.
- **IHB geo**: Energía de interacción en la base IHB<sup>?</sup> entre las moléculas con optimización geométrica.
- **d inter**: Distancias de puente de hidrógeno que hay en las bases S66 e IHB. En el apéndice A están indicadas cuales son.
- **d intra**: Distancias de enlace en la base COV.

## 3.2. Resultados

### 3.2.1. C, H, O, N

Para empezar a optimizar la base se utilizaron los elementos mas abundantes en biomoléculas, H, C, N y O.

La optimización de las NAOS para estos elementos se hizo de forma iterativa. Se empezó con unos valores iniciales de carga atómica  $q_{\alpha\mu}$ , radio de corte  $R_c(\alpha\mu)$  para el C, N y O. En el caso del hidrógeno al solo tener un electrón se optimizo la base con el 'mixing factor', ( $c_H$ ).



Para estas iteraciones se variaba uno de los parámetro, ya sea carga o radio de corte, de un elemento y los demás parámetros se dejaban fijos. Esto nos daba unas curvas de variación para cada parámetro donde elegimos el valor que minimizara los diferentes MAD en cada elemento. Una vez escogido el valor hay que hacer de nuevo otra iteración ya que aunque variábamos de forma independiente los parámetros realmente al cambiar uno afectaba a los demás y había que trabajar de forma iterativa. De esta forma fuimos buscando la mejor base que tuviera en cuenta tanto la energía de interacción entre moléculas como las distancias intra-moleculares e inter-moleculares.

Para minimizar el número de variables para el C, N y O el radio de corte para el orbital 's',  $R_c(s)$ , se hizo que siempre fuera un poco más pequeño que el del orbital 'p',  $R_c(p)$ , para el mismo átomo.

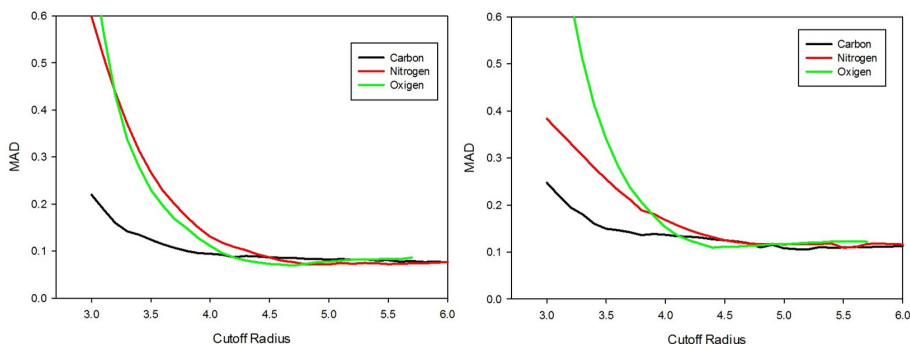
$$R_c(s) = R_c(p) - 0,5 \quad (3.4)$$

Las distancias están en unidades atómicas (u.a.). En el caso de la carga se mantuvo la carga del orbital 's' siempre en  $q_s = 2,0$  y solo se variaba la carga del orbital 'p',  $q_p$ .

Los valores finales de los parámetros  $q_{\alpha\mu}$ ,  $R_c(\alpha\mu)$  y  $c_H$  se escogieron de forma que tuvieran un buen compromiso para las diferentes medidas, en concreto tuvimos especial interes en las distancias y energías de los puentes de hidrógeno.

El valor que se obtuvo para los diferentes parámetro después de todas las iteraciones fue:

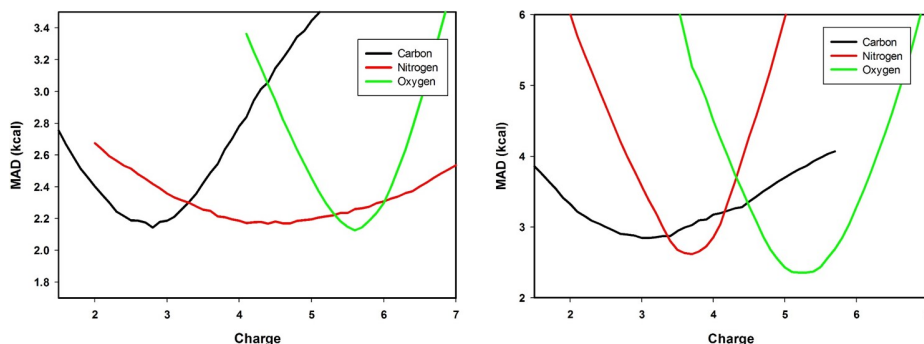
- Hidrógeno,  $Rc_s = 4.0$  u.a.,  $\chi_s(r) = 0.8$
- Carbono,  $Rc_s = 4.0$  u.a.,  $q_s = 2.0$ ,  $Rc_p = 4.5$  u.a.,  $q_p = 1.0$
- Nitrógeno,  $Rc_s = 4.0$  u.a.,  $q_s = 2.0$ ,  $Rc_p = 4.5$  u.a.,  $q_p = 2.0$
- Oxígeno,  $Rc_s = 3.7$  u.a.,  $q_s = 2.0$ ,  $Rc_p = 4.2$  u.a.,  $q_p = 3.8$



**Figura 3.2** Desviación absoluta media (MAD) para las energías de interacción intermolecular en función del  $R_c(p)$  para los átomos C, N y O. (a) S66 fix<sup>2</sup> para moléculas neutras con interacciones no covalentes. (b) IHB fix<sup>2</sup> Mean absolute deviation (MAD) para moléculas cargadas con interacciones no covalentes. En esta figura los radios de corte son  $R_c(s) = R_c(p) - 0,5$  para el elemento que se está optimizando y los demás parámetros para el resto de átomos son los de la base optimizada ( $q_{\alpha\mu}, c_H, R_c(H)$ ).

En las figuras 3.2, 3.3 y 3.4 se ven diferentes ejemplos de el proceso de optimización. En la figura 3.2 se compara la energía de interacción inter-molecular como función del radio de corte  $R_c(p)$  para los elementos C, N y O para los 'benchmarks' S66 e IHB sin optimización geométrica, S66 fix e IHB fix. Las figuras con la energía de interacción del S66 e IHB con optimización geométrica están en el apéndice A. En estas figuras se ve como la base mejora al hacer los radios de corte mas largos. El MAD en este caso tiene un decaimiento exponencial al aumenta  $R_c$ , para mejorar nuestra eficiencia computacional los radios de corte fueron elegidos para ser lo más pequeños posibles sin llegar a afectar a la calidad de la base. Los resultados elegidos fueron:  $R_c(p) = 4.5, 4.5$  y  $4.2$  a.u. para C, N y O respectivamente. Para el H se hizo un análisis similar y se obtuvo  $R_c = 4.0$  a.u. (ver apéndice A).

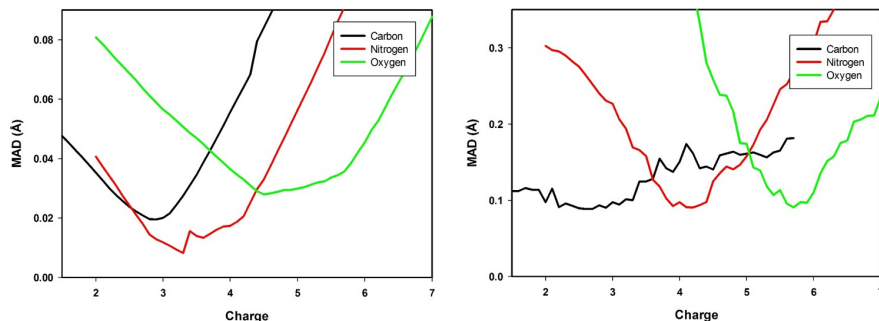
Una vez elegido el valor de  $R_c$  lo suficientemente largo se generaron diferente NAOs para cada elemento cambiando la carga atómica  $q_\mu$ , ver figura 3.1. En la figura 3.3 se representa la comparación de las energías



**Figura 3.3** Desviación absoluta media (MAD) para las energías de interacción intermolecular en función de la carga total  $Q = q_s + q_p$  utilizada para definir los orbitales de los átomos de C, N y O. (a) S66 fix<sup>2</sup> para moléculas neutras con interacciones no covalentes. (b) IHB fix<sup>2</sup> Mean absolute deviation (MAD) para moléculas cargadas con interacciones no covalentes. En esta figura la carga para el elemento optimizado la carga del orbital 's' es  $q_s = 2,0$  y los demás parámetros para el resto de átomos son los de la base optimizada ( $q_{\alpha\mu}, c_H, R_c(H)$ ).

de interacción inter-moleculares con los 'benchmarks' S66 e IHB al variar la carga total  $Q = q_s + q_p$  de los elementos C, N y O sin minimizar las estructuras (las gráficas con minimización están en el apéndice A). En estos resultados se como el O tiene una variación muy clara en función de  $Q$  con un mínimo en la región  $Q \sim 5,4 - 5,8$  tanto en el S66 fix como en el IHB fix. En el nitrógeno hay un mínimo claro para el conjunto IHB fix en  $Q \sim 4,0$ , en cambio para el S66 fix el mínimo es mucho mas ancho y esta distribuido en valores de  $Q \sim 4,0 - 5,5$ . Por último el C es bastante plano en ambos S66 fix e IHB fix y tiene un mínimo en  $Q \sim 3,0 - 3,5$ . esto se debe a que la mayoría de las interacciones se producen entre átomos que tienen N y O, ya que es una base pensada para biología, lo que hace las variaciones en el carbono no afecten demasiado a estas energías de interacción.

En la figura 3.4 se muestra una comparación de las distancias intra-moleculares ( $d_{intra}$ ) e inter-moleculares ( $d_{inter}$ ) al variar la carga  $Q$  para los distintos elementos. En ambas gráficas podemos ver como existen mínimos bien marcados para todos los elementos. Para el C se observa



**Figura 3.4** Desviación absoluta media (MAD) para las distancias (a) intra-moleculares ( $d_{\text{intra}}$ ) y (b) inter-moleculares ( $d_{\text{inter}}$ ) en función de la carga total  $Q = q_s + q_p$  utilizada para definir los orbitales de los átomos de C, N y O comparandolas con los valores experimentales y de los 'benchmarks'. En esta figura la carga para el elemento optimizado la carga del orbital 's' es  $q_s = 2,0$  y los demás parámetros para el resto de átomos son los de la base optimizada ( $q_{\alpha\mu}, c_H, R_c(H)$ ).

que los mínimos de los dos gráficas tienen un valor muy parecido  $Q \sim 2,9$ . Sin embargo para el N y el O los valores para  $d_{\text{intra}}$  nos piden cargas más pequeñas  $Q \sim 3,3$  y  $Q \sim 4,6$  respectivamente que para las  $d_{\text{inter}}$   $Q \sim 4,1$  y  $Q \sim 5,8$ .

Finalmente obtuvimos un MAD para el S66 de 1.99 kcal/mol que es un resultado excepcional. Si comparamos este resultado con otros métodos como PBE (2.31 kcal/mol), B3LYP (3.49 kcal/mol), PM6 (3.41 kcal/mol), AM1 (6.83 kcal/mol) o SCC-DFTB (3.55 kcal/mol)<sup>42</sup> se puede apreciar que esta base tiene una precisión realmente buena. Además se ha conseguido mantener una base mínima con sol orbitales 's' para el hidrógeno y 's' y 'p' para C, N y O consiguiendo una gran eficiencia computacional.

### 3.2.2. S, P, Mg

A parte de H, C, N y O hay otros elementos que tienen un papel importante en biología, para nuestra base también hemos añadido azufre, fósforo y magnesio. El azufre es un elemento que aparece en algunos

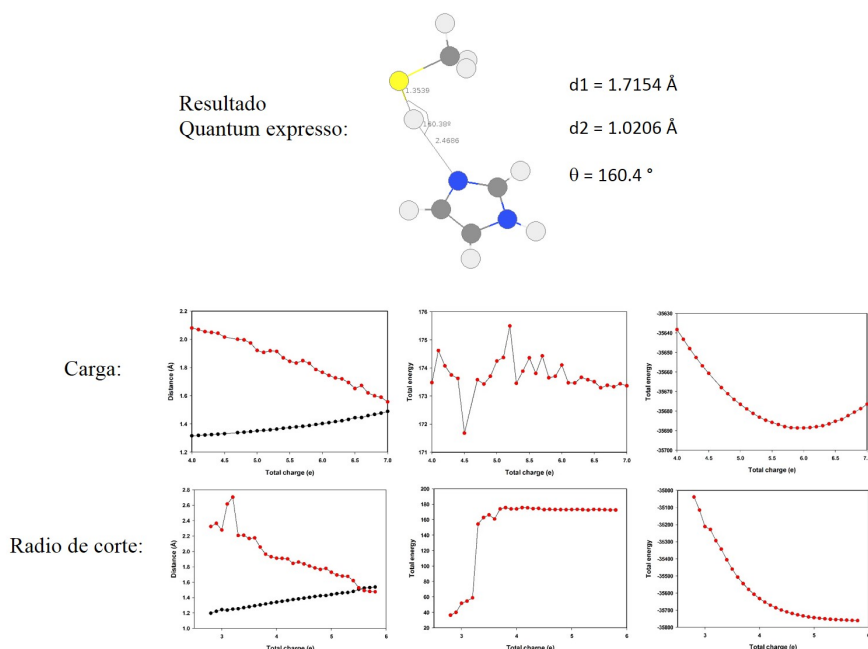
residuos de las proteínas y forma parte del centro catalítico de algunas enzimas. El fósforo es un elemento esencial en biología ya que forma parte de las moléculas de ADN y ARN, también se utiliza para el transporte de energía en el ATP, en las fosforilaciones y otros procesos biológicos. Además el ATP y el resto de nucleótidos suelen estar acompañados de un ion magnesio que es necesario en las reacciones enzimáticas.

Para estos átomos a diferencia de los anteriores no hay bases de datos estandarizadas con las que comparar. Por eso lo que hicimos fue comparar las geometrías y energías de algunos sistemas con cálculos obtenidos con Quantum Espresso, un DFT de ondas planas, con el mismo funcional que utilizamos en fireball (BLYP).

Para el azufre se estudio la geometría para distintos sistemas con enlaces parecidos a los que hay en proteínas. Los sistemas fueron: dimetil sulfuro ( $\text{CH}_3\text{-S-CH}_3$ ) para estudiar el enlace con los carbonos, un puente disulfuro formado por ( $\text{CH}_3\text{-S-S-CH}_3$ ), un metanotiol ( $\text{CH}_4\text{S}$ ) y la interacción de un metanotiol con un imidazol ( $\text{C}_3\text{H}_4\text{N}_2$   $\text{CH}_4\text{S}$ ) y con un ácido acético ( $\text{C}_2\text{H}_4\text{O}_2$   $\text{CH}_4\text{S}$ ). Para el azufre se estudio la inclusión de orbitales 'd' y se vio que el efecto de mejora sobre la base era pequeño para el coste computacional que añadía.

Para el fósforo las moléculas que se utilizaron fueron dos óxidos de fósforo distintos ( $\text{P}_4\text{O}_6$  y  $\text{P}_4\text{O}_{10}$ ), ácido fosfórico ( $\text{PO}_4\text{H}_3$ ) y la interacción del ácido fosfórico con una molécula de agua ( $\text{PO}_4\text{H}_3$   $\text{H}_2\text{O}$ ) y con un amoníaco ( $\text{PO}_4\text{H}_3$   $\text{NH}_3$ ). Para el fósforo si que se vio que el orbital 'd' era totalmente necesario ya que si no no se reproducían las geometrías conocidas.

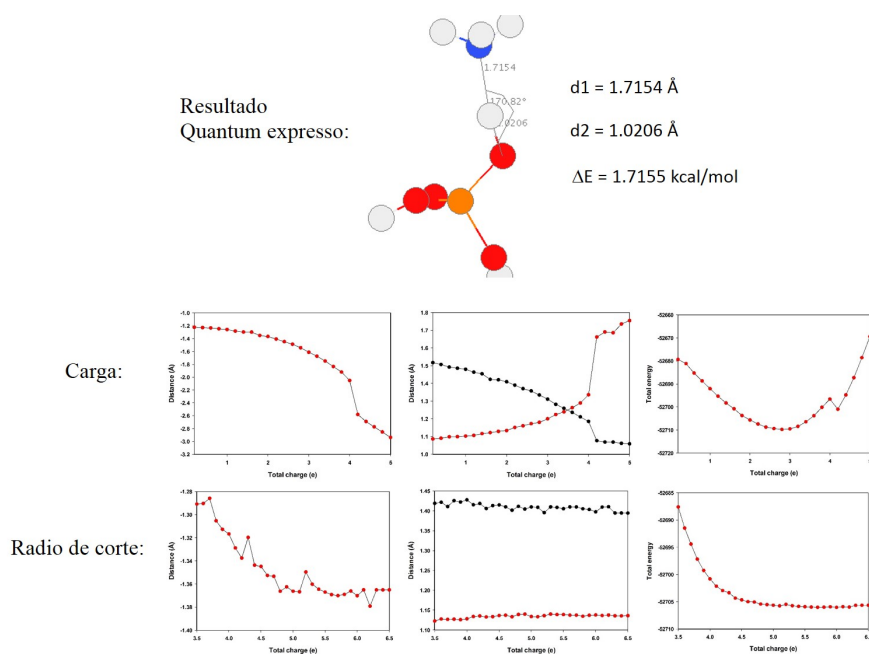
Tanto en el azufre como en el fósforo el comportamiento al variar el radio de corte fue igual que en los elementos de la sección anterior, el valor variaba poco a poco hasta que se se estabiliza para radios suficientemente grandes. Los valores seleccionados para el radio de corte se hizo igual que



**Figura 3.5** Ejemplo de estructura utilizada para optimizar el azufre, metanotiol e imidazol (a) Resultado de la optimización con Quantum expresso. (b) Graficas variando la carga total 'Q' del azufre. (c) Graficas variando el  $R_c(p)$  del azufre.

para los elementos anteriores, lo mas cortos posibles sin que afectara a la calidad de la base.

Para la optimización de las cargas el azufre y el fósforo resultaron ser algo mas problemáticos ya que para algunos valores de carga se cambiaba la protonación en los sistemas que tenían interacción con otras moléculas. Estos casos tuvieron una importancia mayor para optimizar en la optimización de la base. Para optimizar se calculo se compararon los resultados con Quantum Expresso para las distancias de puente de hidrógeno y energías de interacción, también se compararon las distancias de enlace para todas las moléculas.



**Figura 3.6** Ejemplo de estructura utilizada para optimizar el f3sforo, 3cido fosf3rico y amoniaco. (a) Resultado de la optimizaci3n con Quantum expreso. (b) Graficas variando la carga total 'Q' del f3sforo. (c) Graficas variando el  $R_c(p)$  del f3sforo.

En las figuras 3.5 y 3.6 se puede ver dos ejemplos de las medidas que se tenían en cuenta al optimizar el azufre y el fósforo. El resto de estructuras que se han tenido en cuenta están en el apéndice A.

El Magnesio fue un caso diferente, para este elemento se intento reproducir el comportamiento de un ion magnesio 2+. Para optimizar los parámetros se estudio como era su esfera de hidratación cuando estaba rodeado de aguas. Para minimizar la cantidad de carga que cogía el Mg se trabajo con carga total  $Q = 0$  y se probo con con orbitales 's', 'p' y 'd'. Al final se utilizo solo orbitales 's' ya que se podia reproducir solo con estos el comportamiento de Quantum Espresso y además de esta forma se minimizaba la cantidad de carga Lowding para el átomo de Mg.

Los resultados de la optimización de estos elementos fue:

- Magnesio,  $Rc_s = 4.5$  u.a.,  $q_s = 0.0$
- Azufre,  $Rc_s = 3.8$  u.a.,  $q_s = 2.0$ ,  $Rc_p = 4.3$  u.a.,  $q_p = 3.5$
- Fósforo,  $Rc_s = 4.5$  u.a.,  $q_s = 1.0$ ,  $Rc_p = 5.0$  u.a.,  $q_p = 1.4$ ,  $Rc_d = 7.0$  u.a.,  $q_d = 0.0$

### 3.3. Test in QM/MM simulations

Una vez optimizada la base se probó en distintos centros activos de proteínas para estudiar su estabilidad. Estas pruebas se hicieron en dinámicas sin restricciones donde la región de QM era el centro activo de las enzimas. Para ello se buscaron enzimas que tuvieran los distintos átomos en su centro activo. Para probar la base con H, C, N y O se utiliza la enzima tripsina donde se vio que los átomos de la triada catalítica permanecía estable y en una geometría muy parecida a la de dinámica clásica. Para la base con azufre se probo en enzima CRAT y para la el fósforo en el centro activo de F1FO. En ambos casos también se vio la estabilidad y que no había ningún cambio en la protonación de los residuos.



### 3.4. Conclusiones

Se ha optimizado una base de orbitales numéricos de tipo atómico para el método DFT FIREBALL para su uso en biomoléculas. En esta optimización se ha tenido en cuenta diferentes 'benchmark' típicos de sistemas biológicos para el hidrógeno, carbono, nitrógeno y oxígeno consiguiendo un MAD de 1.99 kcal/mol para el 'benchmark' S66 con las estructuras fijas. Además gracias a que la base se ha mantenido con la mínima cantidad de orbitales posibles la eficiencia computacional FIREBALL va a ser muy alta. Esto nos permitirá hacer simulaciones de dinámica molecular QM/MM largas en sistemas biológicos y poder hacer estudios de energía libre. Esta base tiene los elementos típicos de orgánica (H, C, N y O) que permiten estudiar la mayoría de biomoléculas pero también se han añadido el S, P, Mg que permiten estudiar una gran cantidad de reacciones importantes en biología como el ciclo del azufre o todas las ATPasas que necesitan P y Mg.



## CAPÍTULO 4

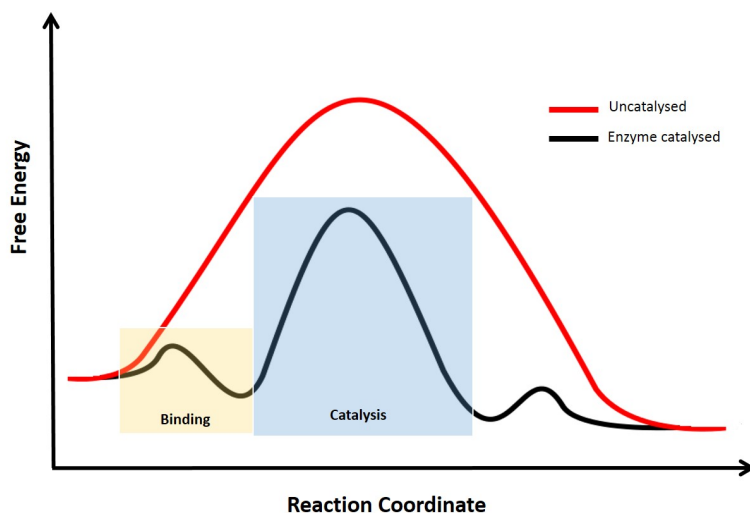
### ESTUDIO DE REACCIONES ENZIMÁTICAS

---

El uso más habitual de los métodos QM/MM es el estudio de reacciones enzimáticas, en este capítulo se van a estudiar dos reacciones enzimáticas distintas con el método de QM/MM desarrollado en el capítulo 2. Para estudiar estos mecanismos se necesitan técnicas de QM/MM debido a que en las reacciones hay grandes cambios en la estructura electrónica del sistema, como puede ser la formación o ruptura de un enlace, que obliga a utilizar técnicas de QM para poder simularlos<sup>11,12</sup>. Por otro lado, el tamaño de estos sistema hace imposible que la utilización técnicas de QM para todo el sistema por el coste computacional y hay que utilizar MM en el resto del sistema.<sup>43</sup>

#### 4.1. Enzimas

Los seres vivos necesitan tener control sobre las reacciones químicas que suceden en su interior. Esto lo consiguen controlando estas reacciones a través de enzimas. Las enzimas son proteínas o ribozimas capaces de modificar la cinética química de las reacciones. Estas modificaciones se producen gracias a la interacción de algunos grupos funcionales de la



**Figura 4.1** Esquema del perfil de energía típico en una reacción enzimática.

enzima con el sustrato. Estas interacciones son las que colocan el sustrato en la posición idónea de forma que se pueda producir la catálisis.

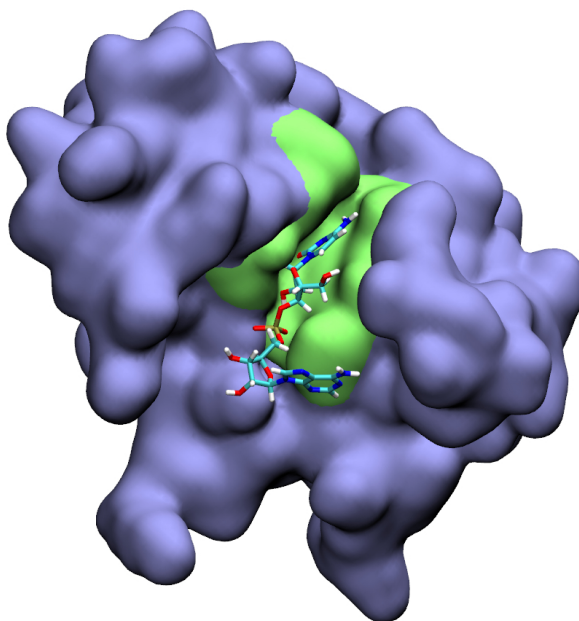
#### 4.1.1. Unión del sustrato

El primer paso en la catálisis enzimática es la unión al sustrato a la proteína, ver figura 4.1. En este proceso el sustrato se coloca a una distancia y orientación adecuada para que se pueda producir la reacción.

Esta unión suele estar mediada por puentes de hidrógeno entre la enzima y el sustrato consiguiendo una gran especificidad. Para entender como se forma esta unión Emil Fischer en 1894 propuso el modelo de 'llave y cerradura' donde la enzima y el sustrato eran modelos rígidos y la especificidad dependía de lo bien que encajaran ambos. Posteriormente en 1958 Daniel Koshland propuso el modelo 'Induced fit' que es el que se utiliza actualmente en el cual tanto la enzima como el sustrato al interactuar el uno con el otro cambiaban su estructura para poder mejorar la interacción entre ambos.

Estas uniones se han estudiado computacionalmente ya que son vitales a la hora de diseñar nuevos fármacos. Los métodos computacionales mas típicos son el 'Docking' y la dinámica molecular. Los métodos de 'Docking' utilizan campos de fuerza más sencillos para calcular la interacción, donde la proteína típicamente está fija o tiene una movilidad muy baja. Estas aproximaciones permiten calcular una gran cantidad de sustratos distintos de forma rápida y sirven para tener un punto de partida sobre qué sustratos encajan y cómo se colocan. Esta unión también se puede estudiar utilizando dinámica molecular clásica, aunque no permite el estudio de una gran cantidad de sustratos distintos, sirve para hacer un análisis mas exhaustivo de la interacción. Para poder estudiar esta interacción suelen ser necesarias dinámicas relativamente largas (sobre 20-50 ns) en las que se estudia qué tipo de interacciones forma el sustrato con la enzima y la estabilidad que tienen. También se puede hacer un análisis mas completo de la energía de interacción utilizando un postprocesado de la dinámica con MMPBSA<sup>44</sup> o utilizar QM/MM para tener un análisis mas exacto de las interacciones<sup>45</sup>.

Estudiar cómo es la unión del sustrato a la enzima es el primer paso que hay que hacer al estudiar una reacción enzimática con QM/MM. Este estudio nos dará información sobre la estabilidad de las interacciones y también podremos hacernos una idea de qué átomos va a ser necesario estudiar con QM. Para ello se hace una dinámica clásica libre de unos 10-20 ns y se comprueba que el sistema está estable, tanto la estructura general de la proteína como las interacciones del centro activo. A continuación también hay que estudiar esa estabilidad con una simulación QM/MM donde se comprueba que los átomos que hemos elegido para la región de QM en el sistema sigue siendo estable. Estas dinámicas QM/MM donde se estudia la estabilidad del centro activo no pueden ser tan largas como las de clásica debido al coste computacional, aunque hay que



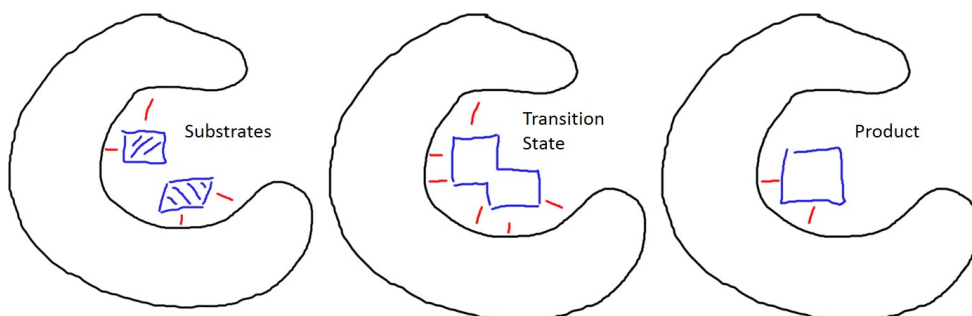
**Figura 4.2** Unión de un dímero de RNA con una enzima (RNasa A), en color lima están destacados los residuos implicados en la unión.

hacerlas suficientemente largas ( $\sim 100ps$ ) de forma que el sistema puede acomodarse al cambio de campo de fuerzas al que le hemos sometido.

Por ejemplo en la figura 4.2 se puede ver la unión del RNA a la RNasa A después de una dinámica clásica y que residuos hay implicados en la unión al sustrato.

#### 4.1.2. Catálisis

El siguiente paso en las reacciones enzimáticas es la catálisis de la reacción, ver figura 4.1. En esta etapa es donde se producen las reacciones químicas y la proteína se encarga de poder cambiar la cinética de estas reacciones. La estructura electrónica del sustrato tiene que cambiar para dar lugar al producto de la reacción. El cambio en la estructura electrónica



**Figura 4.3** Esquema de las interacciones que se producen a lo largo de la reacción. El estado de transición suele ser la estructura con más interacciones para poder catalizar la reacción.

suele tener una barrera energética y la proteína puede influir bajando la altura de esta barrera, catalizando la reacción.

Para disminuir la energía del estado de transición las enzimas pueden utilizar distintas estrategias. Una de las más habituales es crear un ambiente local, con una distribución de cargas específica, de forma que la reacción se produzca con mayor facilidad. También se puede disminuir la barrera del estado de transición formando intermediarios covalentes entre proteína y sustrato de forma cambiando el camino de la reacción. A la hora de simular estas reacciones es importantísimo tener en cuenta todas estas posibilidades que existen para la catálisis y además que nuestro método QM/MM sea capaz de modelizar como el ambiente (que típicamente estará simulado en clásica) influye en la región activa (que suele estar simulado con QM). En el capítulo 2 hemos visto como nuestro método QM/MM es capaz de reproducir este efecto en el ejemplo del puente salino.

Además este ambiente suele ser especialmente importante en el estado de transición ya que la cantidad de interacciones con la proteína se suele maximizar para poder reducir la barrera, ver figura 4.3.

En 1989 Svante Arrhenius propuso una dependencia de la constante de velocidad de la reacción con la temperatura con la siguiente forma:

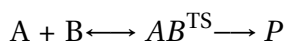
$$k = Ae^{\frac{-E_a}{RT}} \quad (4.1)$$

Donde  $A$  es una constante que depende de cada enzima,  $E_a$  es la energía de activación,  $R$  la constante de los gases y  $T$  la temperatura. Esta relación empírica se ha utilizado para calcular experimentalmente la energía de activación para una gran cantidad de enzimas

Para intentar entender la dependencia de la barrera en una reacción con la constante de velocidad Henry Eyring, Meredith Gwynne Evans y Michael Polanyi en 1935 desarrollaron un marco teórico en la "Teoría del estado de transición". La teoría del estado de transición asume dos hipótesis:

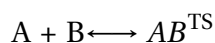
- Cuasi equilibrio entre los reactivos y el estado de transición.
- Una vez pasado el estado de transición se forma el producto directamente y la reacción es irreversible.

Lo que da lugar a una reacción con la siguiente forma:



donde  $A$  y  $B$  son el sustrato,  $AB^{TS}$  es el estado de transición y  $P$  es el producto.

Podemos dividir la reacción en dos partes, la primera sería entre los reactivos y el estado de transición:



Al suponer que están en cuasi equilibrio la constante de equilibrio  $K^+$  se puede escribir en función de sus concentraciones

$$K^{\dagger} = \frac{[AB]^{TS}}{[A][B]} \quad (4.2)$$

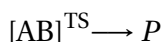


y aplicando la mecánica estadística en esta ecuación llegamos a:

$$K^\dagger = \frac{k_b T}{h\nu} e^{\frac{-\Delta G^{TS}}{RT}} \quad (4.3)$$

donde la constante de equilibrio entre los reactivos y el estado de transición depende de la diferencia de energía libre entre los dos estados,  $\Delta G^{TS}$ , la temperatura,  $T$ , y la frecuencia de vibración del estado de transición,  $\nu$ . La constante de reacción para llegar del estado de transición al producto que depende del modo vibracional del estado de transición.

La segunda parte de la reacción estudia el paso del estado de transición al producto. En esta parte los estados no están en equilibrio entre ellos de forma que una vez superado el estado de transición el sistema formara el producto.



y tendremos una ecuación de velocidad para el producto

$$\frac{d[P]}{dt} = k^* \cdot [AB] = k^* \cdot K^\dagger \cdot [A][B] \quad (4.4)$$

donde  $k^*$  es la constante para la segunda parte reacción y  $k$  la constante para toda la reacción.

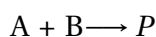
$$k = k^* \cdot K^\dagger \quad (4.5)$$

y  $k^*$  es

$$k^* = \nu \cdot \kappa \quad (4.6)$$

donde  $\nu$  es la frecuencia de vibración del estado de transición y  $\kappa$  es una constante distinta para cada reacción.

Si juntamos estas dos partes tendríamos la constante de reacción para la ecuación de Eyring:



$$k = \kappa \frac{k_b T}{h} e^{\frac{-\Delta G^{TS}}{RT}} \quad (4.7)$$

Y vemos que la forma de la ecuación tiene también un factor exponencial

$$k \propto e^{\frac{-\Delta G^{TS}}{RT}} \quad (4.8)$$

como en la ecuación de Arrhenius

$$k \propto e^{\frac{-E_a}{RT}}. \quad (4.9)$$

Por lo que podemos asumir que la energía de activación que se calcula experimentalmente con la ecuación de Arrhenius es equivalente a la energía del estado de transición  $\Delta G^{TS}$ . Esta energía libre del estado de transición se puede calcular computacionalmente y entender que interacciones producen una bajada en la energía de la barrera.

En la teoría del estado de transición se hacen muchas aproximaciones y no es válida para todos los sistemas<sup>46</sup>. Por ello se ha seguido trabajando sobre ella y se han ido añadiendo modificaciones. Estas modificaciones se deben a cosas como la probabilidad de efecto túnel en los hidrógenos<sup>47</sup>, la forma del estado de transición 'Variational transition state'<sup>48</sup>, que haya varios caminos de energía parecida 'Multi-path variational transition state theory'<sup>49</sup>, ...

En la mayoría de los casos el principal factor del cual va a depender la constante de reacción,  $k$ , va a ser la energía del estado de transición. Por ello estudiar cual es la geometría de este estado nos aporta una gran cantidad de información a cerca de como se produce la catálisis. Durante los últimos años se han estudiado computacionalmente cómo se producen las reacciones de muchas enzimas, pero en la gran mayoría de estudios lo que se calcula son perfiles de energía potencial o se busca solo la geometría y energía del estado de transición<sup>50</sup> cuando la constante de reacción realmente depende de la energía libre. Esto se debe a que los

cálculos de QM no son demasiado efectivos computacionalmente y el estudio dinámico de las reacciones requiere el cálculo de un conjunto de estructuras. En el caso de fireball, debido a su eficiencia computacional, sí que nos permite utilizar métodos de energía libre para poder ver el perfil de la barrera. Para poder reproducir bien el ambiente enzimático hay que utilizar métodos QM/MM y en la actualidad prácticamente todos los estudios de reacciones enzimáticas se realizan con QM/MM<sup>8,51,52</sup> aunque no hay demasiados estudios con espacios conformacionales amplios<sup>53</sup> o utilizando dinámica<sup>54</sup>.

### 4.1.3. Salida del producto

Una vez terminada la catálisis, el producto de la reacción todavía esta unido a la enzima. Para poder salir, normalmente tiene que pasar una pequeña barrera energética. Esta barrera, cuando existe, suele estar relacionada con un pequeño cambio conformacional y muchas veces en proteínas motores esta salida, al soltarse la tensión, es la que desencadena los cambios conformacionales.

## 4.2. Estudio dinámico de reacciones en biomoléculas

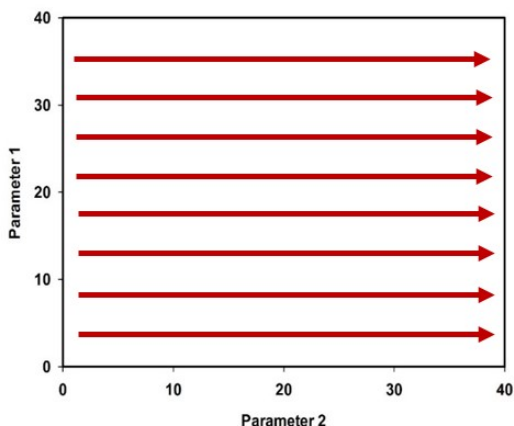
Los sistemas biomoleculares tienen una gran cantidad de átomos y trabajan a temperatura fisiológica. Por tanto estos sistemas tienen una gran cantidad de grados de libertad y para un estado macroscópico hay muchas configuraciones posibles. Este gran número de estados posibles se puede tener en cuenta haciendo Dinámicas Moleculares (MD). Si queremos estudiar el conjunto de configuraciones posibles a lo largo de una reacción enzimática hay que hacer dinámicas largas y utilizar QM/MM. Estas dinámicas largas son prohibitivas computacionalmente

para la mayoría de métodos QM/MM, por ello la mayoría de estudios de reacciones enzimáticas todavía se hace con minimizaciones. El problema de este tipo de estudios es que al no tener en cuenta la temperatura cada estado está representado únicamente por una estructura<sup>8</sup>.

Para poder analizar una reacción biomolecular hay que determinar las coordenadas de reacción que están relacionadas con los cambios que se van a producir en la reacción. Cada valor de estas coordenadas va a definir un estado diferente del sistema que va a corresponder a un conjunto de estructuras microscópicas distintas en lugar de una estructura atómica única para cada estado. Al tener para cada estado un conjunto de estructuras y energías se puede estudiar la variación de energía libre a lo largo de las coordenadas de reacción. Para estudiar estas variaciones se han propuesto una gran cantidad de metodologías, WHAM, thermodynamic integration, metadynamics, free energy perturbation, ...<sup>55-59</sup>

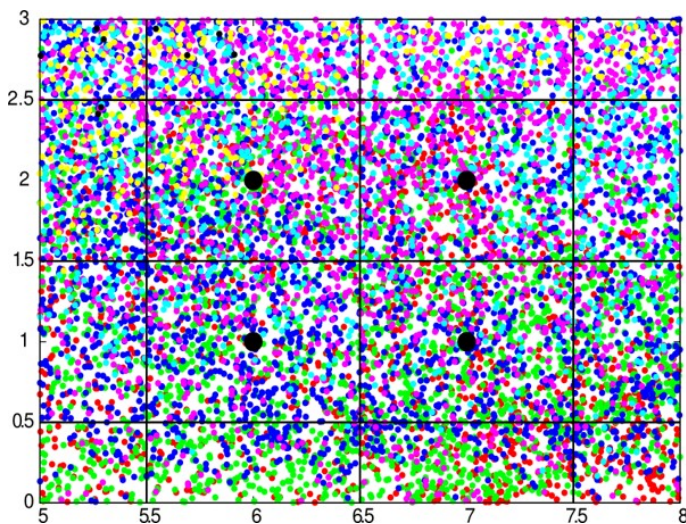
Al estudiar reacciones enzimáticas con QM/MM la dinámica a RT puede tener un papel muy importante aunque normalmente no se toman en consideración debido al alto coste computacional de los métodos QM para hacer dinámicas y explorar un amplio espacio conformacional. Para solucionar el problema del coste computacional se pueden usar métodos semi-empíricos como AM1 y PM3<sup>60</sup>, aunque tienen una precisión menor que va a afectar sobre todo a la energía de la barrera. La implementación FIREBALL/AMBER que hemos desarrollado nos permite estudiar grandes espacios conformacionales con una precisión mayor y un coste computacional similar a los semi-empíricos usando DFT.

Para calcular la energía libre de las reacciones utilizamos una aproximación en la que típicamente vamos a tener dos coordenadas de reacción. Por ejemplo, una reacción enzimática normalmente puede ser caracterizada por dos coordenadas de reacción, las distancias de los enlaces covalentes que se van a romper y crear. Después, para explorar el espacio conformacional, se utilizaron métodos de dinámica



**Figura 4.4** Esquema de como se explora el espacio conformacional a traves de distintos 'steered'.

molecular dirigida como el 'steered' o el 'umbrella sampling'. De esta forma generamos muchas de dinámicas moleculares y se intenta explorar un espacio conformacional delimitado de forma homogénea. Al explorar esta espacio conformacional hay que variar las coordenadas del sistema de una forma lo suficientemente suave para que el sistema pueda tener los cambios de forma adiabática. Normalmente esto la hacemos a través de 'steered' suficientemente lentos en una coordenada y la otra se mantiene fija, ver figura 4.4. De forma que si generamos muchas trayectorias con diferentes valores equiespaciados de esta coordenada fija se puede generar una distribución homogénea y densa de estructuras en nuestro espacio conformacional. Estos mapas de energía suelen tener alrededor de  $2 \times 10^6$  estructuras individuales cada una con unos valores de coordenadas de reacción y energía, ver figura 4.5.



**Figura 4.5** Zoom de una parte de la grid, los puntos grandes negros son los puntos de la gris y las líneas delimitan que zona pertenece a ese punto. Cada punto pequeño de colores una estructura distinta con su correspondiente energía, el color indica que es la misma trayectoria.

Posteriormente, generamos una grid uniforme de puntos en el plano definido por las dos coordenadas de reacción y se le asigna cada una de las estructuras al punto más cercano de la grid. De esta forma cada punto de la grid esta asociado a un conjunto de estructuras atómicas; normalmente estos puntos de la grid contienen  $\sim 10^3$  estructuras distintas. A partir de esto el mapa de energía libre se calcula de la siguiente forma.

Primero se calcula el valor de la función  $Q$  para cada punto de  $(x, y)$  la grid

$$Q(x, y) = \frac{1}{N_i(x, y)} \sum_i \exp\left(\frac{-E_i(x, y)}{k_B T}\right); \quad (4.10)$$

donde  $x$  e  $y$  son los valores de las coordenadas de reacción,  $E_i(x, y)$  es a energía potencial para todas las estructuras atómicas que corresponden al punto  $(x, y)$  de la grid y  $N_i(x, y)$  es el número de estructuras de la suma. Y

la energía libre se obtiene con

$$F(x, y) = -k_B T \ln [Q(x, y)]. \quad (4.11)$$

Finalmente para obtener el mapa se aplica un modo de regresión local en 3D (LOESS) para reducir las irregularidades inherentes a la recogida de datos estadística,<sup>61</sup> y a que las variaciones de energía libre en una superficie tienen que ser suaves.

En este método estamos suponiendo que la distribución de los estados no depende de la dinámica intrínseca del sistema sino que está homogéneamente distribuida ya que movemos nosotros con 'steered' las coordenadas de reacción y por eso la energía se pesa a través de la exponencial. También suponemos que con una cantidad relativamente corta de datos por punto de la grid  $\sim 10^3$  es suficiente para poder calcular la energía libre del sistema. Ambos puntos son bastante difíciles de conseguir y por eso esta forma de calcular a veces puede llevar a tener bastante ruido y hay que hacer un suavizado (LOESS) para que los datos tengan algo de sentido. El punto mas complicado de conseguir es la cantidad de estructuras necesarias por punto de la grid ya que a veces las variaciones de energía potencial son muy grandes y se necesitaría ir mucho mas lento en los 'steered' para que el sistema visite una cantidad de estados accesibles cercanos a la grid mucho mayor. Esto provoca en general que la exactitud de la energía libre calculada no sea demasiado alta.

Por otro lado este método tiene la ventaja de que al depender de la energía de los puntos no depende del camino por el que se llega a ese estado, ya que se calcula de forma independiente cada punto de la grid, y por eso no arrastra errores al pasar por zonas que no están demasiado bien muestreadas, ver comparación con WHAM en el capítulo 5 para la dimerización de timina. También te permite hacerte una idea rápida y sin un muestreo muy grande cual va a ser el comportamiento dinámico del sistema.

## 4.3. RNasa A

### 4.3.1. Introducción

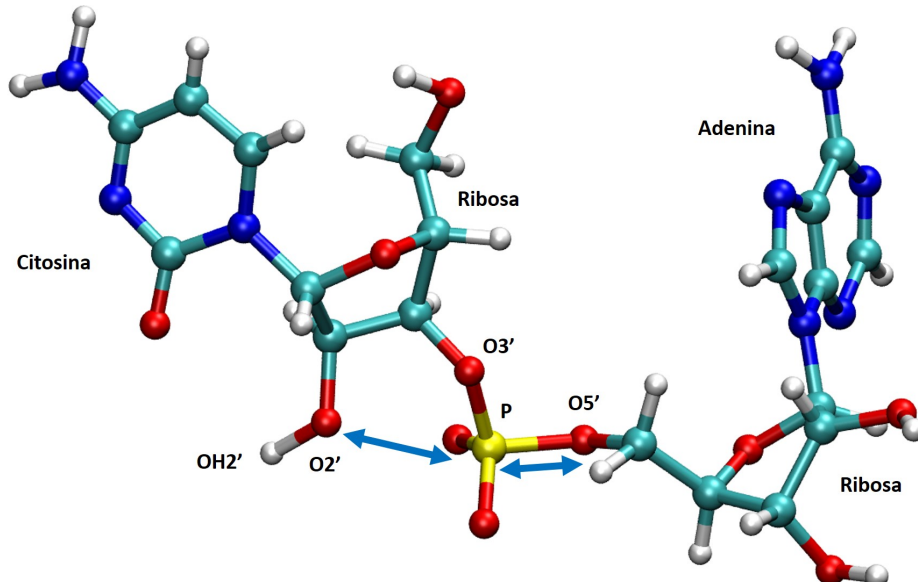
La primera reacción enzimática en la que estudiaremos el mecanismo de reacción con el método FIREBALL/AMBER es la escisión de la molécula de ARN a través de una reacción de hidrólisis que está catalizada por la proteína RNasa A. Las Ribonucleasas o RNasas son enzimas muy abundante ya que la catalizan hidrólisis de ARN cuando ya no se le necesita. Hay una gran cantidad de RNasa distintas para distintos procesos celulares. La Rnasa A hidroliza ARN de cadenas simples y no tiene ningún tipo de especificidad. Además es una de las mas utilizadas en los laboratorios para degradar ARN ya que aguanta temperaturas muy altas. Esta enzima fue la primera proteína en la cual se describió el mecanismo de reacción en el año 1961.<sup>62</sup> Esta ruptura se produce en dos etapas, en la primera etapa el enlace fosfodiester del ARN se rompe terminando en un nucleótido cíclico 2', 3'. En la segunda etapa este nucleótido cíclico se hidroliza y termina en un nucleotido 3'. Nosotros vamos a estudiar computacionalmente la primera etapa de esta reacción.

El mecanismo molecular por el que este proceso se lleva a cabo esta mediado por dos residuos de histidina, His 12 y His 119 que se desprotonan y protonan respectivamente al inicio de la reacción, ver figura 4.6 y 4.7. En esta reacción la His 12 actúa como un aceptor del protón del O2' del azúcar y la his 119 esta envuelta en la transferencia del protón al O5'.<sup>63</sup>

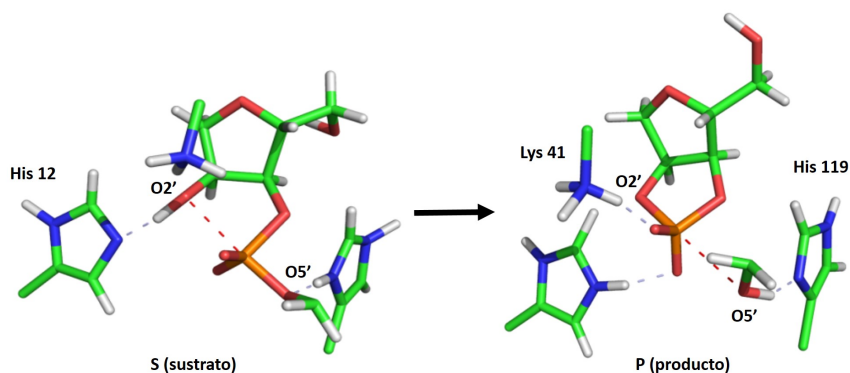
### 4.3.2. Resultados

Como coordenadas iniciales para la simulación se utilizó la estructura de rayos X de la RNasa A con código 1RPG del 'Protein Data Bank'; las coordenadas de inicio del sustrato (dímero de ARN) fueron obtenidas por semejanza con el inhibidor presente en el cristal.<sup>64</sup> Se solvató el sistema

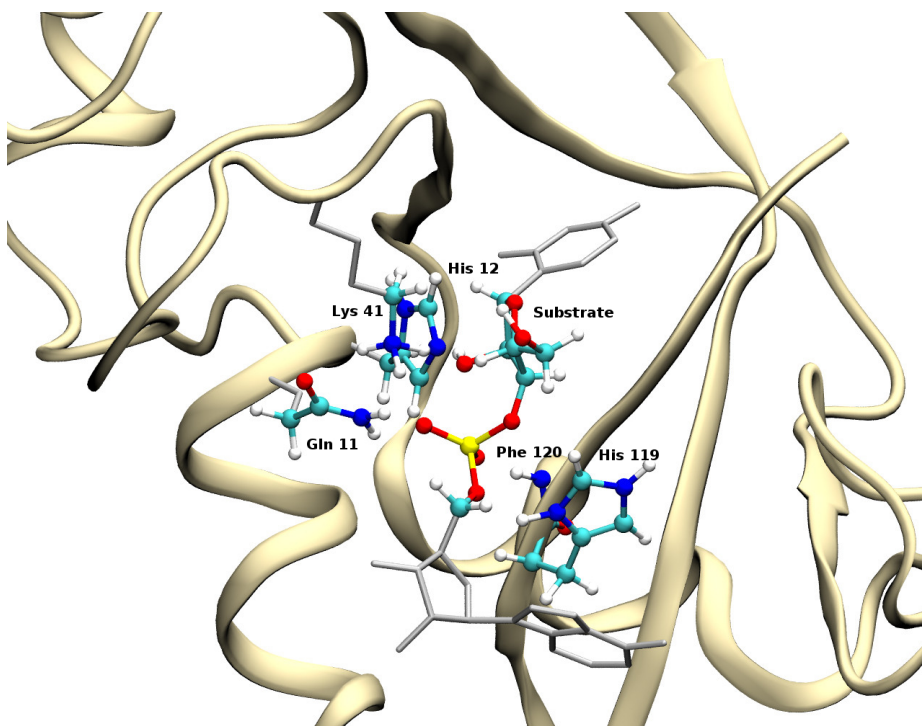




**Figura 4.6** Dímero de ARN con los nombres de los átomos que van a tener un papel relevante en la reacción. En flechas azules están indicadas las coordenadas de reacción



**Figura 4.7** Estructuras inicial (izquierda) y final (derecha) de la primera etapa de la reacción de escisión del ARN catalizada por la RNasa A. También están indicadas los residuos que juegan un papel importante en la reacción



**Figura 4.8** Centro activo de la RNasa A. La región de QM esta representada en 'ball and sticks', el resto de la proteína en 'cartoons' amarillos y el ácido nucleico en 'sticks' grises. Las aguas y los iones no se han representado para mayor claridad.

añadiendo aguas TIP3P y se neutralizó con iones de  $Cl^-$ . Antes de empezar la simulación QM/MM se hizo una dinámica clásica sin restricciones de 10 ns, con el campo de fuerza ff99SB<sup>65</sup> y un cut off de 10Å, donde se vio que el sistema permanecía estable durante toda la simulación. Después se continuó con una dinámica de 100 ps de QM/MM MD con una región de QM que incluía el azúcar con el OH2', el enlace fosfodiéster, las cadenas laterales de los residuos His12, His119, Lys41, Gln11 y la cadena principal de la Phe120 (figura 4.8), en total se incluyeron 75 átomos en la región de QM

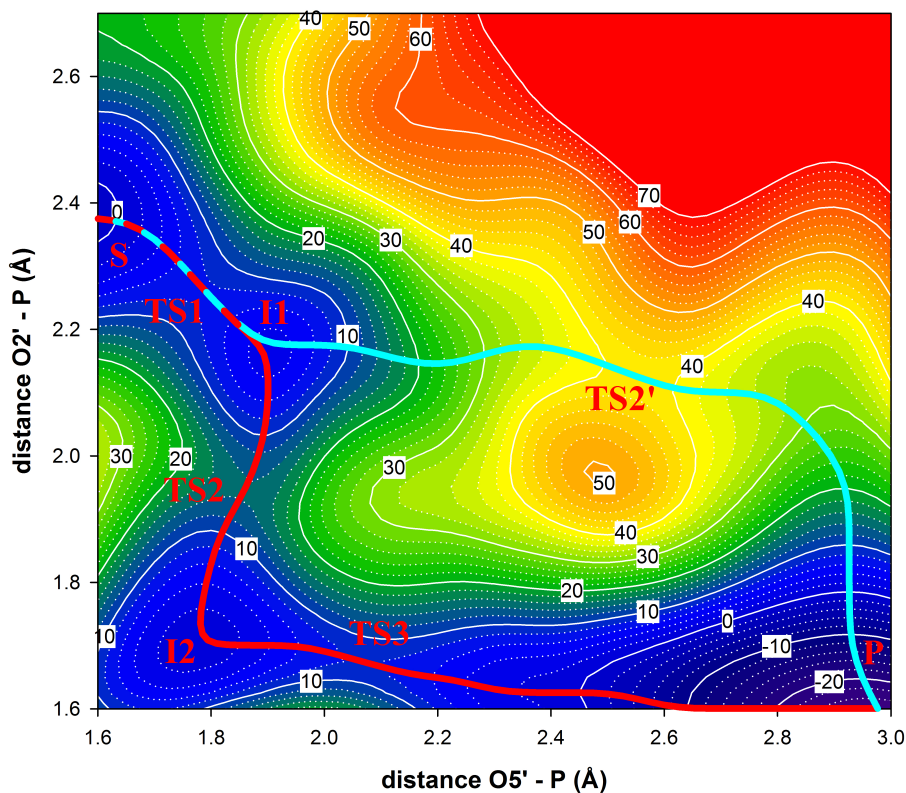
Para intentar entender en profundidad el primer paso de la reacción de hidrólisis de la RNase A calculamos un mapa de energía libre con nuestra

implementación QM/MM (ver capítulo 3) utilizando como coordenadas de reacción las distancias O2'-P y O5'-P, ver figura 4.6, que corresponden a los enlaces que se van a formar y romper respectivamente entre O y P.

Para calcular la energía libre se utilizó la misma técnica expuesta en la sección anterior que nos permite explorar un gran espacio conformacional y así poder determinar el camino de mínima energía y la energía del estado de transición. Para obtener este perfil de energía libre (ver imagen 4.9) se utilizaron una gran cantidad de simulaciones QM/MM MD que sumaron un total 2 ns (con un  $\Delta t = 0,5$  fs, *i.e.*  $4 \times 10^6$  steps).

De este mapa de energía libre se puede obtener una gran cantidad de información. La más llamativa es la existencia de dos caminos posibles para la reacción, que están representados en la imagen 4.9. El camino de mínima energía (línea roja en las figuras 4.9 y 4.10) tiene las características de una sustitución nucleofílica biomolecular (mecanismo asociativo o reacción  $S_N2$ ), siendo un camino más complejo donde aparecen dos estados intermediarios (I1 e I2) y tres estados de transición (TS1, TS2 y TS3). El camino alternativo (línea cian en las figuras 4.9 y 4.10) es algo más simple (solo tiene un estado intermediario I1 y dos estados de transición TS1 y TS2') y es compatible con una sustitución nucleofílica unimolecular (mecanismo disociativo o reacción  $S_N1$ ). Este segundo camino al tener una barrera bastante más alta es más desfavorable energicamente y por lo tanto la reacción irá casi siempre por el camino de reacción  $S_N2$ .

Ambos caminos,  $S_N1$  y  $S_N2$ , empiezan con la activación del grupo OH2' de la ribosa donando el protón a la His 12. Esta barrera de energía está dentro del valor típico de una transferencia de protones, alrededor de 8 kcal/mol donde al desprotonarse el O<sup>-</sup>2' la carga negativa del oxígeno está estabilizada por la Lys 41 de la RNasa en el primer estado de transición (TS1). Además el protón transferido a la His 12 la convierte en un residuo positivo que ayuda a estabilizar el estado intermediario I1. Entonces en la reacción asociativa  $S_N2$ , línea roja en las figuras 4.9 y 4.10, el O<sup>-</sup>2'

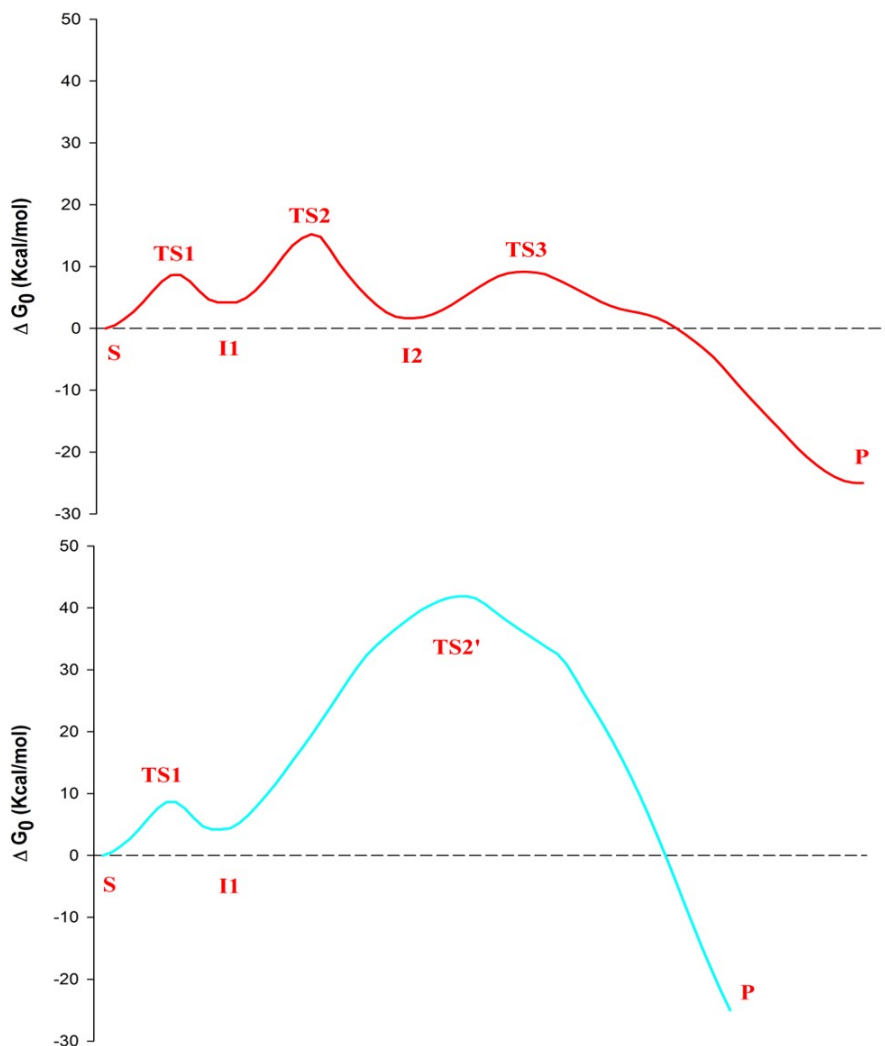


**Figura 4.9** Superficie de energía libre en kcal/mol de la reacción de Rnase A, usando la implementación FIREBALL/AMBER. Los ejes representan las coordenadas de reacción, distancias O2'-P y O5'-P. Se han indicado las dos trayectorias posibles de reacción : disociativa  $S_N1$  en cian y asociativa  $S_N2$  en rojo. También están indicados la posición del sustrato (P), estados de transición (TS), estados intermediarios (I) y el producto de la reacción (P) para cada trayectoria.

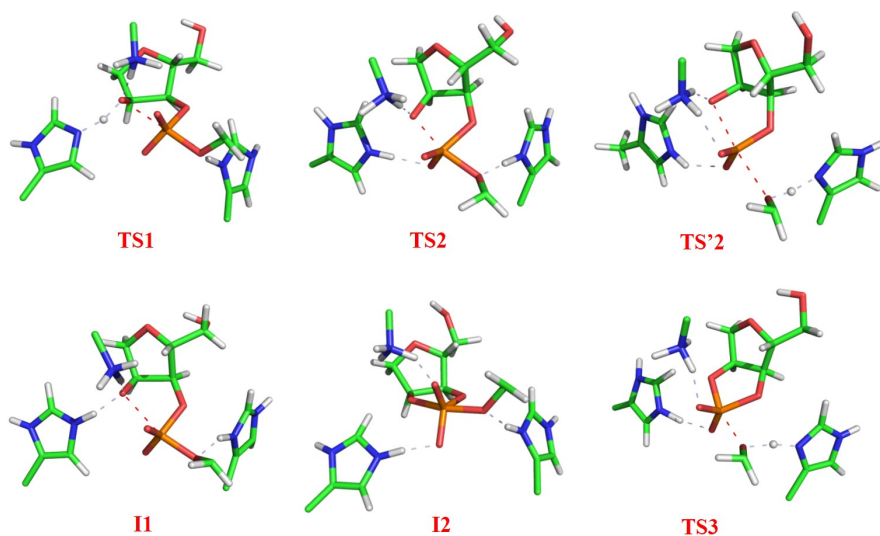
nucleofílico ataca al fósforo del enlace fosfodiéster de forma que se llega al estado de transición (TS2) típico de las reacciones  $S_N2$ , el estado de transición pentacovalente del fósforo, a través de una barrera de 15 kcal/mol. Para estabilizar este estado de transición el imidazol positivo de la His 12 y el grupo amonio de la Lys 41 se mueven hacia el oxígeno ecuatorial del estado pentacovalente del fósforo. Esta estabilización del estado pentacovalente produce una estructura intermedia, I2, que corresponde al intermediario diatómico descrito teóricamente en los estudios de Glennon and Warshel<sup>66</sup>. La reacción finaliza a través del estado de transición TS3 que corresponde a la transferencia del protón de la His 119 al O5' y la formación del nucleótido cíclico 2',3' llegando de esta forma al producto final de la reacción, ver figura 4.7.

El camino disociativo  $S_N1$ , línea cian en las figuras 4.9 y 4.10, comparte el primer paso de la reacción (TS1 y I1) con la reacción  $S_N2$ . En cambio a partir del estado I1 donde el protón ha sido transferido a la His 12 la reacción se produce de forma totalmente distinta. En este caso la His 119 protonada cede el protón a O5' del fosfato mientras se rompe el enlace O5'-P a través del estado de transición TS2' de 40 kcal/mol. El estado de transición TS2' tiene el estado de transición trigonal típico de las reacciones  $S_N1$  y está estabilizado por las interacciones de la His 12 y la Lys 41 con el fosfato y la Lys 41 también estabiliza el O2'. Después del estado de transición TS2' este camino de reacción va directo al producto final sin pasar por ningún otro estado intermedio. Este camino de reacción tiene una barrera de energía mucho mayor que la reacción  $S_N2$  y por lo tanto será mucho menos probable.

Para comparar nuestra técnica con otros métodos existentes se hicieron varias dinámicas QM/MM MD con distintos métodos semi-empíricos. La ventaja que tienen los métodos semiempíricos es que permiten el estudio de espacios conformacionales grandes y cálculos de energía libre debido a su eficiencia computacional pero con una precisión bastante menor. Lo que



**Figura 4.10** Análisis energético de las dos trayectorias posibles para la reacción de la RNasa A obtenidas en el mapa de energía libre 4.9. Reacción disociativa  $S_N1$  en cian y reacción asociativa  $S_N2$  en rojo. También están indicados la posición del sustrato (S), estados de transición (TS), estados intermediarios (I) y el producto de la reacción (P) para cada trayectoria.



**Figura 4.11** Analisis estructural los estados de transición e intermediarios de las dos trayectorias posibles para la reacción de la RNasa A obtenidas en el mapa de energía libre 4.9.

observamos es que estos métodos no eran capaces de describir de forma suficiente mente buena el estado pentacovalente del fósforo y el transporte de protones asociado a el. En algunos casos la transferencia del H del OH2' a la His 12 no se producía bloqueando así la reacción (métodos (AM1<sup>67</sup> y PM3<sup>60</sup> que no tiene orbitales d) o se producía a una distancia demasiado corta del enlace O2'-P y la reacción tampoco conseguía terminar (AM1d<sup>68</sup> o PM6<sup>69</sup>). También se hicieron pruebas con el método DFTB (con la base de átomos con orbitales d en el P)<sup>70</sup> en el cual la estructura inicial se volvía inestable.

### 4.3.3. Discusión

Gracias a la eficiencia computacional de FIREBALL/AMBER hemos sido capaces de explorar un espacio conformacional muy extenso y su energía,

y ver los dos caminos posibles de la primera etapa de la reacción de escisión del RNA haciendo la mínima cantidad de asunciones posibles sobre las trayectorias de la reacción. De esta forma podemos proponer que el camino propuesto  $S_N1$  tiene una energía bastante más alta y que el camino más probable es la reacción  $S_N2$ . Además hemos sido capaces de ver qué residuos estabilizan los estados de transición para ambos caminos y determinar que la diferencia de energía de los dos caminos se produce gracias a que en la reacción asociativa se separa el TS2' de la disociativa en dos barreras mas pequeñas. Esto se produce gracias a que la His 12 y la Lys 41 pueden estabilizar de forma independiente los estados de transición (TS2 y TS3) de forma separada consiguiendo esta reducción en la barrera.

## 4.4. Triosa-fosfato isomerasa

### 4.4.1. Introducción

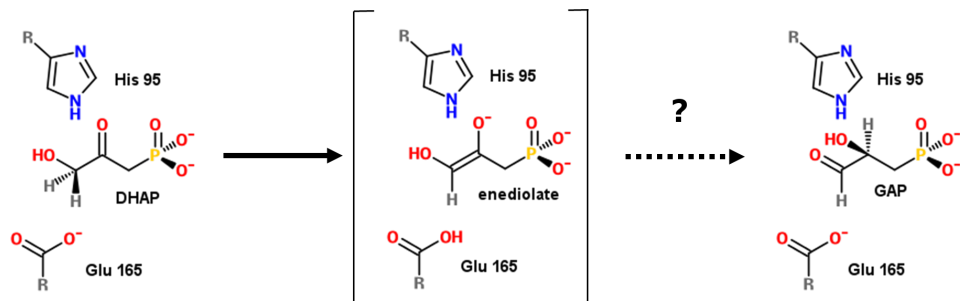
La siguiente reacción estudiada con el método FIREBALL/AMBER es la inter-conversión reversible entre dihidroxiacetona fosfato (DHAP) y gliceraldehido 3-fosfato (GAP) que está catalizada por la proteína Triosa-fosfato isomerasa (TIM). Esta enzima esta implicada en la glucolisis donde tiene un papel fundamental en la eficiencia energética del proceso y por eso esta presente en practicamente todos los seres vivos y es una de las secuencias más conservadas. Esta enzima se ha estudiado de forma extensiva con distintas técnicas tanto experimentales como computacionales, aunque todavía no están claros todos los detalles del mecanismo catalítico. Esta generalmente aceptado que el primer paso de la reacción es el paso del protón del átomo C1 al glutámico 165 dando lugar a un estado intermediario formado por un enediolato, este enediolato se ha detectado también en la reacción no catalizada.<sup>71</sup> En cambio para la



segunda parte de la reacción todavía no está claro como se produce (ver figura 4.12).

Para esta segunda parte de la reacción donde se va del enodialato intermediario al GAP, se han propuesto diferentes caminos posibles para la transferencia del protón del átomo O1 al O2. En uno de los mecanismos propuestos,<sup>72</sup> la His 95 actúa como un ácido catalítico donando el protón al oxígeno del grupo carboxílico en C2, ver figura 4.12. Después el protón del grupo alcoholico O1 es transferido a la His 95 y el C2 recibe el protón del Glu 165 recuperando de esta forma la enzima su configuración original. Los datos de RMN en cambio parecen indicar que el anillo imidazol de la histidina 95 no cambia su carga para cualquier pH dentro del rango de actividad enzimática<sup>73</sup>. Por lo tanto, el camino propuesto implicaría la formación de anión imidazol a pesar del alto valor de pKa ( $\approx 14$ )<sup>74</sup> asociado a este proceso. Para evitar la formación de este anión imidazol se han propuesto otros caminos posibles, como puede ser la transferencia del protón del Glu 165 al oxígeno carboxílico de C2 y la siguiente protonación de C2 desde el grupo alcohol de C1 a traves del Glu 165.<sup>75</sup> Además se ha estudiado la importancia de otros residuos como la Lys 12<sup>76</sup> y el Glu 97<sup>77</sup> que han sido propuestos como candidatos a tener un rol importante en el mecanismo catalítico para evitar la formación del anión imidazol. También se ha propuesto la transferencia directa del protón O1 al O2 sin pasar por ningún residuo pero obtuvieron que la energía era desfavorable si el Glu 165 permanecía protonado.<sup>75</sup>

La incertidumbre sobre el paso del protón puede ser debida a que los perfiles de energía obtenidos para los diferentes caminos se han estudiado a través de técnicas de minimización de energía donde cada estado de protonación del sistema se define *a priori* y se busca el estado de transición entre ellos. El uso de técnicas de MD para calcular un potencial de fuerza media y estudiar el espacio conformacional de la reacción sin especificar ningún estado de protonación específico de los residuos puede ayudar a



**Figura 4.12** Esquema de la conversión de DHAP a GAP catalizada por TIM.

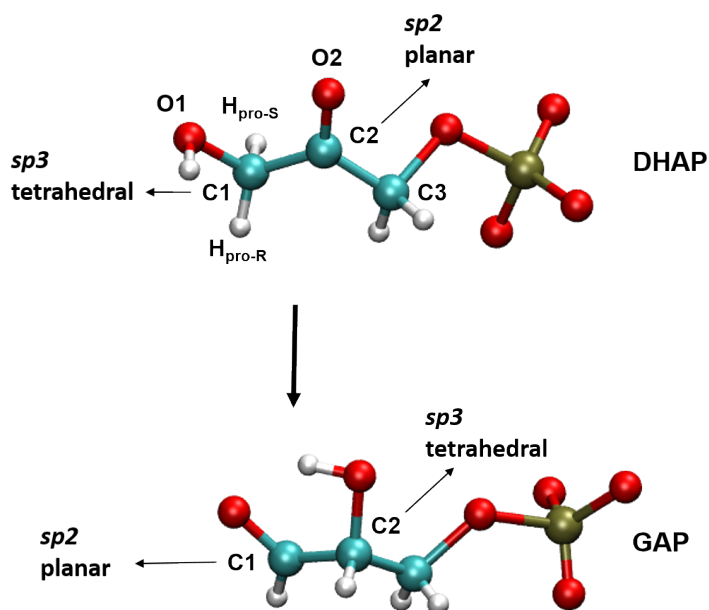
clarificar el mecanismo de transferencia de los protones. Además el uso de técnicas de MD (*e.g.* a  $T=300$  K) puede tener en cuenta algunos efectos dinámicos como la formación de puentes de hidrógeno de barrera baja donde este hidrógeno puede moverse de forma casi libre entre los dos átomos pesados.<sup>78,79</sup>

Durante la reacción catalizada por TIM se puede observar como hay un cambio en la hibridación del carbono C1 que cambia de  $sp^3$  en DHAP a  $sp^2$  en GAP, mientras que la hibridación del carbono C2 va de  $sp^3$  a  $sp^2$ , ver figura 4.13. Esta reorganización de los orbitales modifica la geometría de los enlaces de los átomos C1 y C2 de planar ( $sp^2$ ) a tetraédrica ( $sp^3$ ), y *vice versa* (figura 4.13).

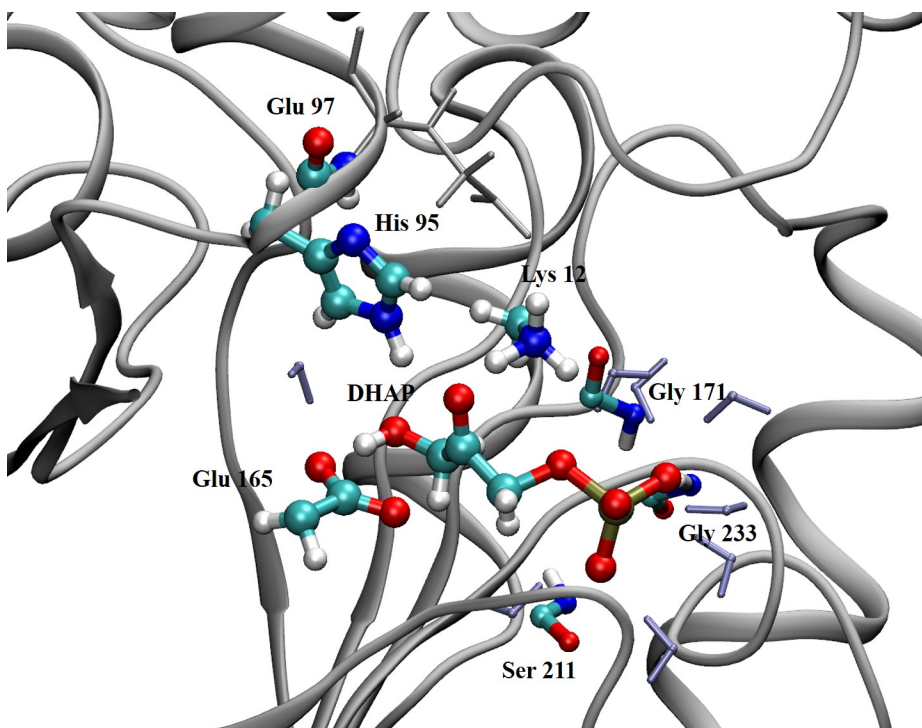
Por tanto, los ángulos diedros de los orbitales C1 y C2 parecen unas buenas coordenadas de reacciones en las que no se hace a priori ninguna suposición sobre los estados de protonación del sistema a lo largo de la reacción.

#### 4.4.2. Resultados

Se estabilizó el sistema con 20 ns de dinámica clásica ( $T = 300$  K) partiendo de la estructura de rayos X de TIM (pdb: 1NEY)<sup>80</sup> y se comprobó que todo seguía estable. Posteriormente se realizó otra dinámica sin restricciones de QM/MM MD de 200 ps con los siguientes átomos en la



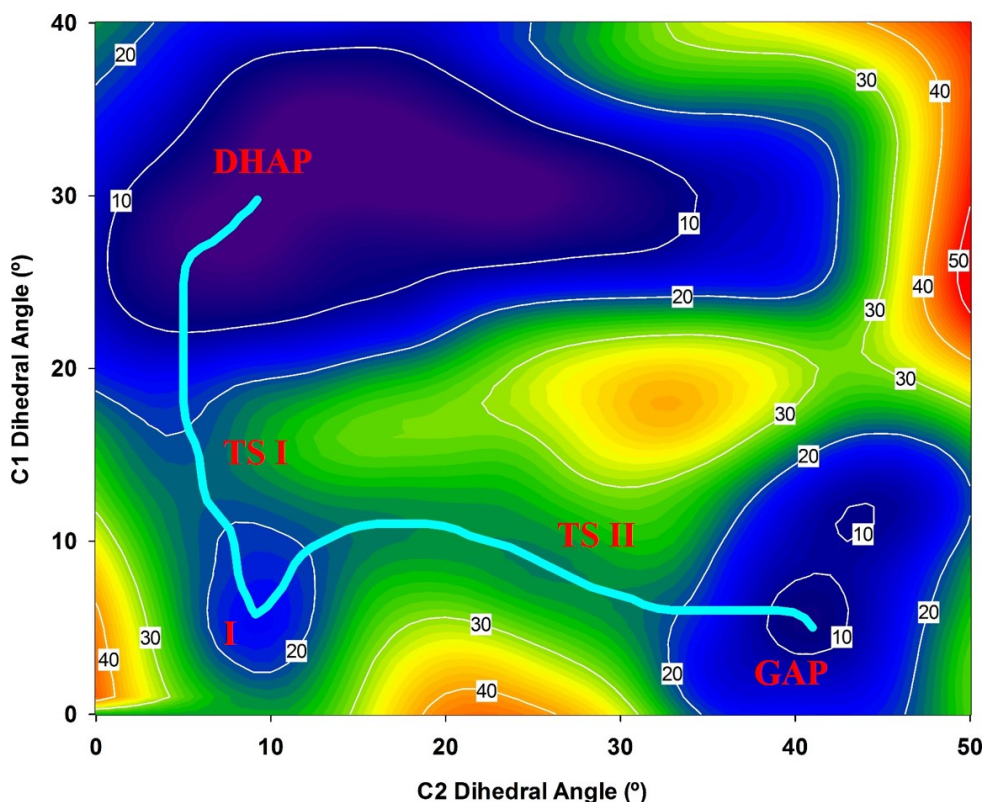
**Figura 4.13** Geometría e hibridación de la dihidroxiacetona fosfato (DHAP) y del gliceraldehído 3-fosfato (GAP). Los átomos involucrados en los dos diedros utilizados como coordenadas de reacción son: O1- C2 - H<sub>pro-S</sub> - C1 para el ángulo diedro C1 y C1 - C3 - O2 - C2 para el ángulo diedro C2



**Figura 4.14** Centro activo de la Triosa-fosfato isomerasa. La región de QM esta representada en 'ball and sticks', el resto de la proteína en 'cartoons' grises. Las aguas y las aguas del centro activo en 'sticks' azules.

región de QM, ver figura 4.14: Sustrato(DHAP); cadenas laterales de los residuos catalíticos Glu 165 y His95; cadena principal del residuo 95 para estabilizar a la His 95; cadena principal de los residuos 171, 211 y 233 para estabilizar el grupo fosfato del DHAP; y cadena lateral de la Lys 12. Este último residuo se añadió a la región de QM debido al efecto dramático que tiene su modificación en el ratio  $K_{cat}/K_m$ .<sup>81</sup> En la figura 4.14 se ve una visión global de la región de QM y la proteína.

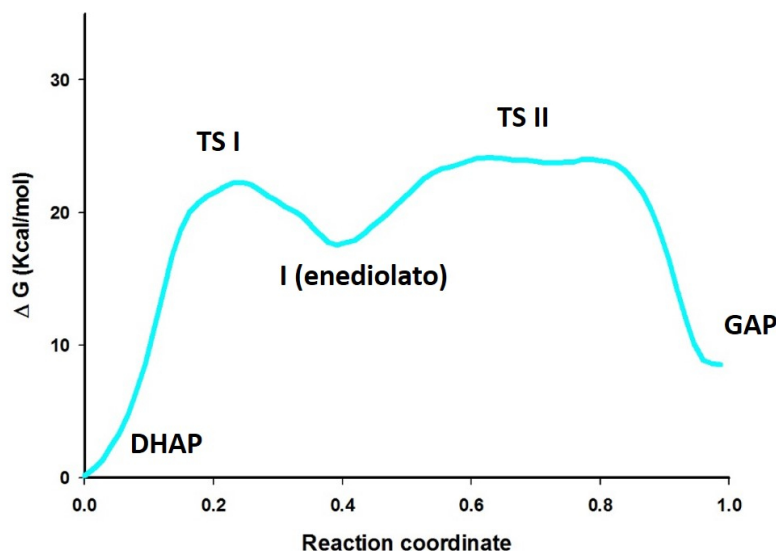
Una vez terminada la dinámica QM/MM MD y con el sistema estable con el campo de fuerza de QM se hizo un mapa de energía libre usando como coordenadas de reacción los dos ángulos diedros C1 y C2 ( $\alpha_1$  y  $\alpha_2$  respectivamente) definidos anteriormente, ver figura 4.13. El mapa se



**Figura 4.15** Mapa de energía libre en kcal/mol de la reacción catalizada por TIM entre DHAP y GAP. Los ejes  $x$  e  $y$  representan las dos coordenadas de reacción, ángulos diedros de los átomos C1 y C2. El camino de mínima energía está señalado con una línea cian.

realizo de la forma expuesta en la sección anterior y se varió  $\alpha_1$  entre 0 y 40 grados y  $\alpha_2$  entre 0 y 50 grados. En la figura 4.15 se pueden ver el mapa de energía libre donde se puede ver los mínimos de energía asociados al sustrato (DHAP), producto (GAP) y al estado intermediario (I).

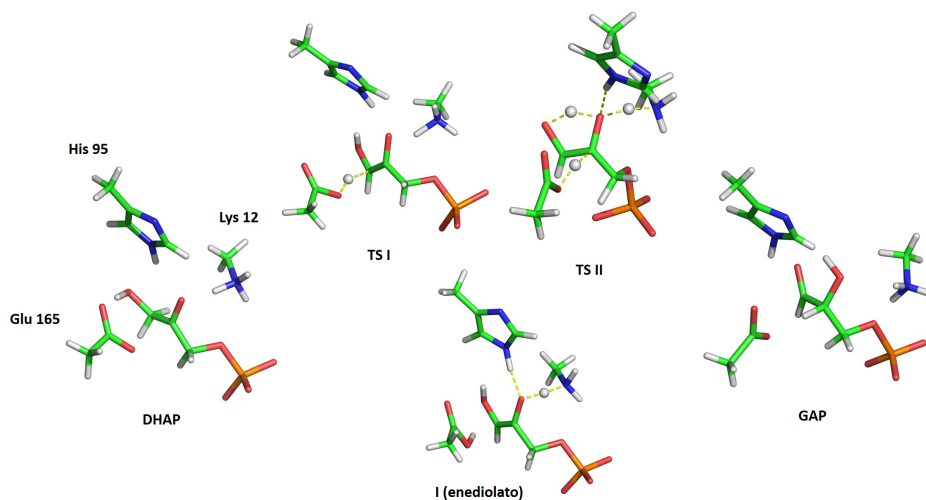
El mínimo de energía mas bajo corresponde al DHAP, el estado inicial de la reacción y se tomó como origen de energía. Este mínimo está centrado en torno a  $\alpha_1 = 30^\circ$  que corresponde a una hibridación  $sp^3$  para el átomo C1. Sin embargo este mínimo es bastante extenso debido a la interacción de algunos residuos con el DHAP (Lys 12 e His 95) que provoca una mayor



**Figura 4.16** Perfil de energía libre para el camino propuesto (línea cian en la figura 4.15) y estructuras atómicas de los estados mas relevantes a lo largo de la trayectoria.

extensión para el ángulo diedro  $\alpha_2$ . El segundo mínimo para los diedros de los átomos C1 y C2 corresponde a una hibridación  $sp^2$  para ambos átomos  $\alpha_2, \alpha_1 \sim 5^\circ - 10^\circ$ . Este mínimo corresponde al estado intermediario (enediolato), ver figura 4.12. El tercer mínimo situado en  $\alpha_1 = 5^\circ$  y  $\alpha_2 = 40^\circ$  corresponden a una hibridación  $sp^2$  para el átomo C1 y una  $sp^3$  para el átomo C2. Estas hibridaciones corresponden al estado final de la reacción (GAP). En la imagen 4.15 también se muestra el camino de mínima energía y por lo tanto el mas probable de la reacción y en la figura 4.16 se ve el perfil de energía de este camino y sus estructuras en la figura 4.17.

La reacción por el camino de mínima energía se puede dividir en dos pasos. En el primer paso va desde el estado inicial al enediolato a través de una barrera de 21 kcal/mol. En otro estudio computacional donde se simulo el paso del sustrato al enediolato utilizando un método de energía libre y QM/MM se obtuvo un resultado muy parecido tanto en energía como en estructura.<sup>82</sup> En el segundo paso de la reacción se pasa de este



**Figura 4.17** Perfil de energía libre para el camino propuesto (línea cian en la figura 4.15) y estructuras atómicas de los estados mas relevantes a lo largo de la trayectoria.

estado intermediario al producto de la reacción (GAP) y pasa a través de un estado de transición con una energía de 23 kcal/mol. Para estudiar los procesos atómicos al detalle a lo largo de este camino se realizó una simulación de QM/MM MD donde se obligó al sistema a ir por el camino de mínima energía a través de un *steered*. En esta dinámica se obligó a los ángulos diedros a moverse por las coordenadas del camino de la figura 4.15 a una velocidad constante para ir del DHAP al GAP en 83000 pasos de dinámica, (ver figura 4.18).

Durante el primer paso de la reacción se produce un reordenamiento de los orbitales del carbono C1 que se puede describir por el cambio de  $\alpha_1$  de  $30^\circ$  ( $sp^3$ ) a  $5^\circ$  ( $sp^2$ ). Durante este proceso el Glu 165 actúa como base catalítica aceptando el  $H_{proR}$  de C1, ver figuras 4.12, 4.16 y 4.17. Después, en el estado intermediario, el enediolato forma un puente de hidrógeno entre el O2 del DHAP y el N $\epsilon$  de la His 95. También se forma otro puente de hidrógeno entre el grupo keto (O2) del DHAP y la Lys 95, además

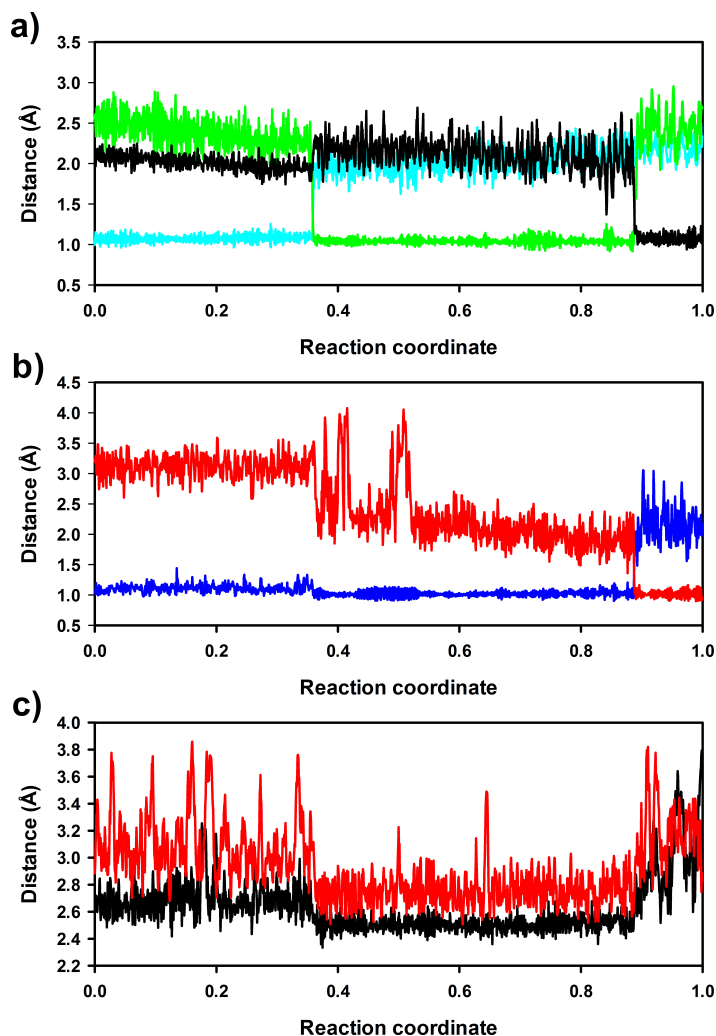
este puente de hidrógeno tiene características de 'low barrier H-bond' y el hidrógeno esta compartido entre el  $N\zeta$  de la lisina y el O2 del DHAP. Estas dos interacciones se encargan de estabilizar enérgicamente el estado intermediario.

En el segundo paso de la reacción el reordenamiento de orbitales se produce en el carbono C2 en el cual vemos como  $\alpha_2$  varia desde  $10^\circ$  ( $sp^2$ ) a  $40^\circ$  ( $sp^3$ ). Durante este proceso el puente de hidrógeno de la His 95 y el 'low barrier H-bond' de la Lys 12 son capaces de estabilizar el aumento de carga negativa del O2 del DHAP que tiene su máximo cuando la reacción pasa por el estado de transición (TS II). En este estado de transición se produce el paso del H del O1 al O2 con una transferencia intramolecular directa y de forma simultanea donde el Glu 165 actúa como ácido catalítico transfiriendo el  $H_{pro-R}$  al C2 del enodiolato y formándose el GAP. Una vez se ha producido la transferencia de protones la interacción entre la His 95 y el O2 se rompe y el 'low barrier H-bond' que formada la Lys 12 vuelve a ser un puente de hidrógeno normal (figura 4.18).

#### 4.4.3. Discusión

El mecanismo de reacción presentado en esta sección presenta algunas características importantes que lo diferencian de los mecanismos propuestos por otros estudios anteriores. La diferencia más notable es la transferencia directa del protón del grupo alcohol al oxígeno carboxílico C2. Esto hace que no sea necesaria la formación de un anión imidazol que parecía bastante desfavorable por los argumentos antes expuestos. También se diferencia en que esta transferencia se produce de forma simultánea a la transferencia del protón del Glu 165 al C2, ver figura 4.18. En estudios anteriores consideraron que la transferencia del protón intramolecular entre O1 y O2 era enérgicamente desfavorable comparada con la transferencia a través del imidazol.<sup>75</sup> Sin embargo estos estudios se hicieron utilizando estrategias de minimización y además el Glu 165





**Figura 4.18** Distancias medidas durante la simulación de QM/MM MD para la trayectoria propuesta (Figura 4.16). Durante la reacción el  $H_{pro-R}$  pasa del C1 al C2 mientras que el otro H va del O1 al O2. (a) distancias entre el  $H_{pro-R}$  y el C1 (cian), distancia entre el  $H_{pro-R}$  y el O del Glu 165 (verde) y distancia entre el  $H_{pro-R}$  y el C2 (negro). (b) distancias entre el H y O1 (azul) y el H y O2 (rojo). (c) distancias de los puentes de hidrógeno entre el O2 y los residuos de alrededor que estabilizan el estado de transición: O2 con N de la His 95 (rojo) y O2 con N de la Lys 12 (negro).

permanecía protonado continuamente. La diferencia de los resultados tiene que ver principalmente por el enfoque utilizado (minimización vs. simulaciones de dinámica), y que en nuestro enfoque no asumimos ningún tipo de estado de protonación de forma que los protones pueden moverse libremente. Esto permite que en nuestra simulación se puedan producir procesos como el paso simultáneo de los dos protones y la formación de un 'low barrier H-bond' entre el O2 la Lys 12 (ver figura 4.18) que estabiliza el estado de transición facilitando la simultaneidad de la transferencia de protones.

## 4.5. Conclusión

En este capítulo hemos presentado una descripción práctica de cómo estudiar reacciones enzimáticas con la implementación de QM/MM( FIREBALL/AMBER ) y usando simulaciones de MD para poder calcular y el camino de reacción y la energía libre asociada a este camino. Esto se ha podido hacer con la combinación de una técnica de DFT de orbitales localizados FIREBALL y los programas de simulación que ofrece AMBER para estudiar sistemas biomoleculares.

El buen balance entre eficiencia computacional y precisión permite estudiar reacciones enzimáticas complejas y poder muestrear un espacio conformacional bastante grande a través de dinámicas QM/MM con una región de cuántica relativamente grande ( $\sim 10^2$  átomos). Gracias a esta eficiencia computacional se han podido hacer estos mapas de energía libre donde se necesita alrededor de  $2 \times 10^6$  estructuras diferentes y con ellas estudiar la energía para los distintos estados del sistema y los posible caminos de las reacciones que queramos estudiar.

Como ejemplos se han estudiado las reacciones de escisión del ARN a través de la reacción catalizada por la RNasa A y la conversión entre dehidroxiacetona fosfato (DHAP) a gliceraldehido

2-fosfato (GAP) que esta catalizada por la triosa-fosfato isomerasa (TIM). Al utilizar la implementación de FIREBALL/AMBER y estudiar un espacio conformacional muy amplio se han podido describir nuevos comportamientos para estas enzimas que no se encontraban en la literatura.



## CAPÍTULO 5

### SIMULACIÓN DE REACCIONES FOTQUÍMICAS EN BIOMOLÉCULAS: DÍMERO DE TIMINA EN ADN

---

La simulación computacional de procesos foto-inducidos en sistemas moleculares grandes como son las biomoléculas es un reto muy complejo. Para poder simular de forma correcta estos procesos hay que poder describir la dinámica del estado excitado de una forma suficientemente correcta y además poder acceder a una gran cantidad de conformaciones para poder explorar el complicado espacio conformacional que tienen estas reacciones. En esta sección se detallará la implementación y el algoritmo que se utiliza en el código DFT de orbitales localizados *Fireball* para simular en QM/MM estados excitados y transiciones electrónicas con dinámica molecular (MDET). Este tipo de procesos son muy comunes en biología donde hay algunos sistemas que son capaces de aprovechar la energía del sol para convertirla en energía química, las proteínas encargadas de la visión, crecimiento de las plantas, foto-lesiones, ..

## 5.1. Dinámica molecular con transiciones electrónicas

Para poder simular excitaciones de luz en moléculas hay mayoritariamente dos tipos de métodos, las dinámicas Ehrenfest (ED) y los métodos de 'surface hopping' (HP).<sup>83-85</sup> En las aproximaciones ED los núcleos se mueven de forma clásica en una superficie de energía potencial (PES) obtenida de una media de todos los estados adiabáticos involucrados. Mientras que en los métodos SH el movimiento de los núcleos proviene de un único estado excitado<sup>83</sup> y hay saltos entre las distintas PES. Los métodos de 'surface hopping' tienen una ventaja sobre lo ED ya que están siempre en un estado adiabático puro y no en un estado no físico media de los estados adiabáticos. Además los métodos ED pueden llegar a estados sin sentido físico, como niveles inaccesibles energeticamente.<sup>86</sup> En el algoritmo de Tully para dinámicas moleculares con transiciones electrónicas (MDET)<sup>87</sup> se utiliza un método de 'surface hopping'<sup>88</sup> para cambiar entre distintos estados. En el método de Tully el sistema evoluciona a través de una PES que corresponde a un determinado estado del sistema y se calcula la probabilidad de transición entre distintos estados con el algoritmo 'fewest -switches'. A continuación se verá como se integran las dinámicas moleculares con transiciones electrónicas en el DFT FIREBALL para poder usarlo con simulaciones QM/MM. El código de dinámica molecular con transiciones electrónicas se divide en los siguientes pasos:

*Paso 1: Solución estacionaria del Hamiltoniano de Kahn-Sham.* En este primer paso se calcula una solución auto-consistente para la configuración atómica actual del hamiltoniano de Kohn-Sham correspondiente a la región de QM.

$$\hat{H}_{KS}[\mathbf{R}]\psi_j(\mathbf{r};\mathbf{R}) = \epsilon_j(\mathbf{R})\psi_j(\mathbf{r};\mathbf{R}), \quad (5.1)$$

Donde  $R$  son las posiciones atómicas en una momento dado de todas las partículas y  $\psi_j(\mathbf{r}; \mathbf{R})$  son los orbitales adiabáticos Kohn-Sham. Los diferentes estados adiabáticos se definen por la ocupación  $f_k$  de cada orbital  $\psi_k$  donde puede ser  $f_k = 1$  (lleno) o 0 (vacío). Por ejemplo para el estado basal se resuelve el determinante de Slater formado por un conjunto de las energías mas bajas de los orbitales de adiabáticos  $\psi_j$  y en el caso de un estado excitado se asume que corresponde a la excitación de un determinante de Slater con un único orbital excitado. De forma que el hamiltoniano de 5.1 se define de forma auto-consistente usando ocupaciones  $f_k$  para cada PES esto se consigue con un 'constrained DFT' o  $\Delta SCF$ <sup>89</sup> en el cual los niveles de ocupación estan impuestos en cada ciclo de auto-consistencia.

*Paso 2: Cálculo de los acoplamientos no adiabáticos.* La evolución temporal de los estados electrónicos se calcula usando la teoría dependiente del tiempo de Kohn-Sham<sup>90-92</sup>, para ello expandimos los orbitales Kohn-Sham dependientes del tiempo  $\varphi_i$  en terminos de los estados Kohn-Sham adiabáticos.

$$\varphi_i(\mathbf{r}, t) = \sum_k c_{ik}(t) \psi_k(\mathbf{r}; \mathbf{R}), \quad (5.2)$$

Estos estados evolucionan temporalmente la siguiente ecuación tipo *Schrödinger*:

$$\hat{H}_{KS}(\mathbf{R}) \varphi_j(\mathbf{r}, t) = i\hbar \frac{\partial \varphi_j(\mathbf{r}, t)}{\partial t}, \quad (5.3)$$

que da lugar a la ecuación de evolución de los coeficientes  $c_{ik}(t)$ :

$$i\hbar \frac{\partial c_{ik}(t)}{\partial t} = c_{ik}(t) \epsilon_k(\mathbf{R}) - i\hbar \sum_m c_{im}(t) \mathbf{d}_{km} \cdot \mathbf{V} \quad (5.4)$$

En esta ecuación el acoplamiento entre el movimiento clásico de los átomos y la evolución de los estados cuánticos electrónicos viene dada por

los términos de acoplamiento no adiabático (NAC):

$$\mathbf{d}_{km} \cdot \mathbf{V} \equiv \sum_{\alpha} \mathbf{d}_{km}^{\alpha} \mathbf{V}_{\alpha} \quad (5.5)$$

donde  $\mathbf{V}_{\alpha}$  es la velocidad del átomo  $\alpha$  y  $\mathbf{d}_{km}^{\alpha}$  son los vectores del acoplamiento no adiabático entre los estados  $k$  y  $m$  para el átomo  $\alpha$  (NACV):

$$\mathbf{d}_{km}^{\alpha} \equiv \langle \psi_k | \frac{\partial \psi_m}{\partial \mathbf{R}_{\alpha}} \rangle \quad (5.6)$$

De forma que la población de los estados adiabáticos  $|C_{km}(t)|^2$  puede sufrir cambios en función de los NACV entre los diferentes estados adiabáticos y las velocidades de los átomos. Los términos  $\mathbf{d}_{km}^{\alpha}$  tienen la ventaja de que en un código de orbitales localizados como *fireball* se pueden calcular en cada paso de computación de una forma bastante eficiente<sup>93</sup>. De forma alternativa también se puede calcular la evolución temporal del NAC, ver ecuación 5.5 de forma numérica<sup>94</sup>:

$$\mathbf{d}_{km} \cdot \mathbf{V} = \langle \psi_k | \frac{\partial \psi_m}{\partial t} \rangle = \frac{1}{2\Delta t} \left[ \langle \psi_k(t - \Delta t) | \psi_m(t) \rangle - \langle \psi_k(t) | \psi_m(t - \Delta t) \rangle \right] \quad (5.7)$$

que da el valor del acoplamiento no adiabático para  $(t - \frac{\Delta t}{2})$ .

*Paso 3: Evolución temporal de los estados electrónicos.* Una vez se han calculado los NAC hay que propagar en el tiempo los coeficientes  $C_{km}$  desde  $(t - \Delta t)$  a  $t$ . para ello se utiliza una integración numérica con el algoritmo *Runge-Kutta*.

*Paso 4: Surface Hopping* En cada paso de simulación  $\Delta t$  existe la posibilidad de que haya una transición entre PES adiabáticos  $E_A \rightarrow E_B$ , la probabilidad de que ocurra este salto se calcula a través del algoritmo *fewest-switches*<sup>87</sup>. En este algoritmo de MDET la probabilidad de que se produzca la transición  $g_{jk}(t)$  esta definida para cada  $\psi_j \rightarrow \psi_k$  en un



intervalo  $[t - \delta t, t]$ :

$$g_{jk}(t) = \frac{b_{kj}(t)\Delta t}{|c_{jj}(t)|^2}, \quad (5.8)$$

donde los coeficientes  $b_{kj}$  estan definidos como:

$$b_{kj}(t) = -2\Re[c_{jk}^*(t)c_{jj}(t)\mathbf{d}_{kj} \cdot \mathbf{V}] \quad (5.9)$$

donde  $\Re$  es la parte real. Para determinar si se produce la transición se utiliza un algoritmo de Monte-Carlo con la probabilidad  $g_{ik}(t)$ . Cuando se produce la transición se cambia la ocupación de los orbitales  $\psi(j)$  y  $\psi(k)$ :  $f_j(1 \rightarrow 0)$  y  $f_k(0 \rightarrow 1)$ , y se vuelven a calcular los estados estacionarios del hamiltoniano de KS de manera auto-consistente con las nuevas ocupaciones  $f_i$ . Para que se conserve la energía se necesita un reescalado de las velocidades de los átomos que se hará en la dirección de los vectores del acoplamiento no adiabático<sup>95,96</sup>, ver ecuación 5.6. En este proceso los estados para subir de nivel electrónico suelen ser denegados ya que tendrían que quitar energía cinética de los átomos y convertirla en potencial, lo cual es bastante difícil. Si este salto no ocurre la simulación continua en el mismo estado electrónico.

*Paso 5: Movimiento de los átomos.* Por último hay que mover los átomos  $R_\alpha(t) \rightarrow R_\alpha(t + \alpha t)$  usando las fuerzas correspondientes a la superficie de energía potencial en la que esté.

$$\mathbf{F}_\alpha = -\frac{\partial E_a}{\partial \mathbf{R}_\alpha} \quad (5.10)$$

y mover los átomos con el algoritmo de integración del movimiento que se utilice.

## 5.2. Simulación de una reacción fotoquímica en ADN: CTD

### 5.2.1. Introducción

La absorción de luz ultravioleta (UV) por el ADN puede dar lugar a diferentes reacciones fotoquímicas que terminan con la formación de enlaces covalentes entre bases adyacentes del ADN<sup>97,98</sup>. El DNA presenta una gran estabilidad frente a las foto-lesiones: alrededor del 99.9% de los fotones absorbidos de la radiación ultravioleta del sol se transforman en calor que se transfiere al solvente sin causar ningún daño<sup>99</sup>. A pesar de que el ozono de la atmósfera absorbe gran parte de la radiación UV nos sigue llegando una gran cantidad de fotones a nuestra piel, y se estima que estos fotones pueden causar unas 50 -100 foto-lesiones por segundo cuando estamos expuestos a la luz solar<sup>100</sup>. Para evitar estas lesiones los organismos han desarrollado diferentes mecanismos como la foto-reactivación, escisión del nucleótido, etc.<sup>101,102</sup>. De las lesiones que se pueden producir al ser absorbido un fotón de luz por el ADN la más abundante es la formación del dímero de ciclobutano en timina (CTD) que tiene como característica la formación de un anillo que incluye dos enlaces C-C entre bases adyacentes<sup>103</sup>.

El desarrollo de nuevas técnicas experimentales ha creado una gran cantidad de información sobre estas reacciones fotoquímicas<sup>97,98,104-111</sup>. Dada la complejidad del DNA la mayoría de los estudios se han hecho sobre pares de base de timina o secuencias sintéticas cortas de oligonucleótidos de ADN de cadena simple en disolución. Se ha visto que en estos sistemas que una vez producida la foto excitación se produce una relajación ultra-rápida de la estructura electrónica al estado basal del sistema. Esta relajación se produce en un tiempo cercano a un picosegundo<sup>106,107</sup>. En particular algunos estudios han señalado que después de la excitación en

cadena simple con todo termina la formación del CTD es una reacción fotoquímica ultrarápida a través del estado singlete que se produce en 1 picosegundo. Con estos resultados se postula que la probabilidad de formación del dímero de timina depende de la conformación del dímero en el momento en el que se absorbe el fotón, y que el bajo rendimiento cuántico se debe a la pocas conformaciones reactivas posibles que hay con el dna a temperatura ambiente (RT) <sup>106,112,113</sup>.

A pesar de estos importantes avances en las técnicas experimentales los estudios teóricos y computacionales todavía son esenciales para poder entender la fotoquímica del ADN <sup>109,114,115</sup>. Debido a la complejidad del sistema la mayoría de los estudios teóricos se han estudiado con las dos nucleobases en estado gaseoso <sup>83,116-122</sup>, usando métodos complejos de química cuántica o con la teoría del funcional de la densidad (DFT). En estos estudios la reacción se lleva a cabo a través de un mecanismo no adiabático en el cual el estado excitado (S1) va sin ninguna barrera a la intersección cónica <sup>123,124</sup> de las superficies de energía potencial (PES) de los estados S1 y S0 que es el embudo para el decaimiento ultra-rápido al estado basal (S0) que da lugar a la foto-lesión.

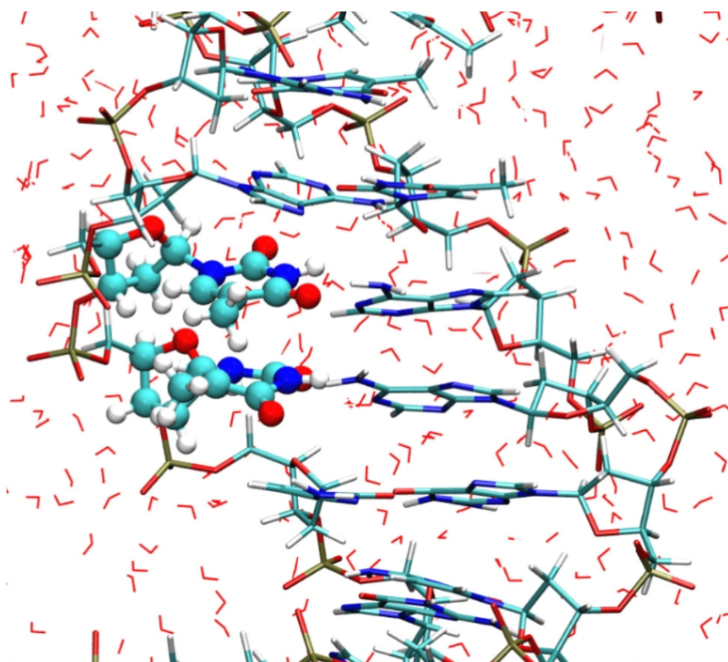
En cambio la dinámica de este proceso debe ser bastante diferente en el caso de un ADN inmerso en el ambiente celular. Estos estudios tampoco tienen en cuenta los efectos dinámicos propios de estar a RT que si se tienen en cuenta al calcular la energía libre. Además para poder simular la dinámica de las reacciones foto químicas hay que ir mas allá de la aproximación de Born-Oppenheimer y tener en cuenta la dinámica del estado excitado y los acoplamientos no adiabáticos de los grados de libertad electrónicos y nucleares. Por todo esto el estudio de las reacciones foto químicas en el ADN es un problema que requiere de la combinación de diferentes estrategias de simulación <sup>114,115</sup>.

En este capítulo analizaremos de forma teórica la formación del CTD en un modelo realista del DNA en condiciones fisiológicas. Para describir

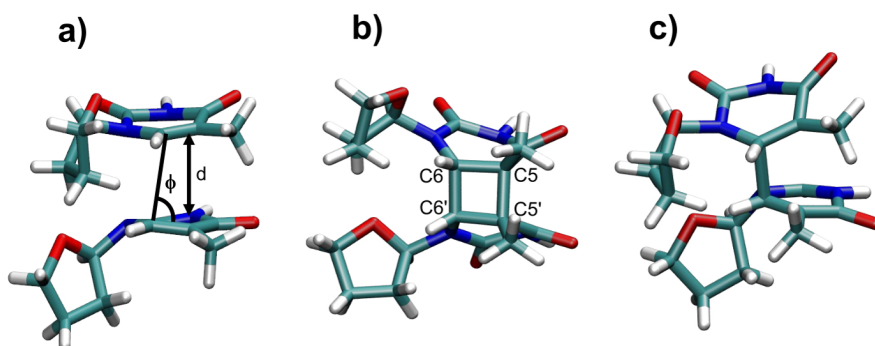
el sistema se utilizara la implementación de QM/MM del capítulo 2, donde se utilizara QM para describir la región donde se produce la reacción foto química mientras que en MM estará el resto de la doble hélice del ADN, las moléculas de agua y los contra iones del solvente<sup>8,15,125</sup>. Primero se determinará el mapa de energía libre<sup>126</sup> para el estado basal, S0, y del primer estado excitado, S1. Estos mapas de energía libre nos permitirán definir, de una forma precisa, la región donde están las *conformaciones reactivas*<sup>106</sup>, i.e. la región del espacio conformacional desde la cual la foto excitación puede provocar que el sistema llegue a formar el CTD en una reacción ultra-rápida. Esta región está relacionada con la *zona de intersección cónica* (CIZ) S1/S0 en la cual se produce la des-excitación de S1 a S0. Posteriormente se realizarán dinámicas no adiabáticas<sup>87,124,127</sup> con QM/MM para analizar de una forma mas detallada el mecanismo atómico de la reacción. Con estas simulaciones se podrá estimar la probabilidad relativa de formación del CTD o la conversión interna al estado estándar del DNA, una vez el sistema ha alcanzado la zona de intersección cónica. Por último se realizaran dinámicas clásicas de MD para calcular la probabilidad que tiene el sistema de estar en la zona reactiva en el momento de la foto excitación. Toda esta información nos dará una imagen detallada del mecanismo de formación del CTD y de la foto-estabilidad del DNA.

### 5.2.2. Método

Nuestro sistema consiste en un fragmento de DNA-B con 10 pares de bases y una esfera de aguas de 30 Å con los contra-iones correspondientes de Na<sup>+</sup>. La secuencia del ADN es CGAATTAAGC y proviene de una estructura cristalográfica con el dímero formado<sup>128</sup>. La región de QM se compone de 52 átomos: las dos bases de timina adyacentes con sus correspondientes desoxirribosas, ver figura 5.1. Las dos timinas están en la misma hebra del DNA y están localizados en el medio del DNA. El resto de átomos del DNA de las dos cadenas (fosfatos y resto de nucleótidos) y



**Figura 5.1** Vista de la parte central de la doble hélice de DNA con la región de QM, el DNA restante y el solvente. La región de QM, representada con '*ball and sticks*', consiste en dos nucleósidos de timina adyacentes en una secuencia de DNA con 10 pares de bases, el resto del DNA y el solvente son modelados con un campo de fuerza clásico.



**Figura 5.2** Representación atomística de las dos bases de timina. (a) DNA estándar; (C) Dímero de ciclobutano en timina; (c) Zona de intersección cónica. Los átomos de carbono C5, C6 (y C5', C6' en la otra base) están indicados en (b). Las coordenadas de reacción utilizadas en (a),  $d$ , es el punto medio de los enlaces C5-C6 de ambas bases y  $\phi$  el ángulo formado por los átomos C5'-C6'-C6.

todas las aguas y contra-iones ( $\text{Na}^+$ ) están incluidos en la región de MM, en total esta región contiene  $\sim 11.000$  átomos.

Las figuras 5.2(a-b) muestran las estructuras atómicas típicas para dos nucleósidos de timina antes y después de la reacción fotoquímica: los dos enlaces dobles  $\text{C5}=\text{C6}$  y  $\text{C5}'=\text{C6}'$  de las timinas adyacentes dan lugar al anillo de ciclobutano al formarse los dos enlaces covalentes  $\text{C5}-\text{C5}'$  y  $\text{C6}-\text{C6}'$  (ver notación en la fig. 5.2b). La figura 5.2a muestra las dos coordenadas de reacción utilizadas en este estudio: la distancia,  $d$ , entre el punto medio de los enlaces C5-C6 de ambas bases y  $\phi$  el ángulo formado por los átomos C5'-C6'-C6. Con estas coordenadas de reacción se ha calculado los mapas de energía libre para los estados electrónicos S1 y S0.

En los análisis computacionales de las reacciones fotoquímicas si se quiere tener en cuenta los grados de libertad y la gran cantidad de estructuras accesibles para un estado a RT se necesita un coste computacional muy alto<sup>114</sup>, por eso se suele ignorar y se asume una única estructura de mínima energía que representa a cada estado. En este

estudio vamos a utilizar el método QM/MM desarrollado en el capítulo 2 FIREBALL/AMBER para calcular los mapas de energía libre de los estados S1 y S0 y además realizar 2560 trayectorias no adiabáticas para estudiar el sistema de forma dinámica.

Para la dinámica clásica se utilizó el programa Amber<sup>129</sup> con el campo de fuerzas *parm99bsc0* para moléculas de DNA, y aguas TIP3P para el solvente. Para los cálculos de DFT se usó un funcional de canje-correlación BLYP<sup>27,28</sup> con unos pseudo-potenciales 'norm-conserving'<sup>130</sup>. La base de orbitales localizados utilizada para el cálculo es la desarrollada en el capítulo 3.

Los mapas de energía libre de las imágenes 5.4, 5.5 y 5.6 fueron obtenidos a través de simulaciones largas de QM/MM MD a temperatura ambiente para las PES de S0 y S1 de la forma indicada en el capítulo 4. Para generar estos mapas se utilizaron  $\sim 3 \times 10^6$  estructuras distintas para cada mapa para poder describir el espacio conformacional del sistema. Las coordenadas de reacción de estos mapas son la distancia  $d$  y el ángulo  $\phi$  comentados anteriormente y variaron entre 1.3 - 5 Å ( $d$ ) y 40° - 120° ( $\phi$ ). Los mapas se han obtenido con una grid de 3000 puntos, con 75 valores distintos de la distancia  $d$  (con pasos de  $\Delta d = 0.05$  Å) y 41 valores para el ángulo  $\phi$  ( $\Delta \phi = 2^\circ$ ). Para cada uno de estos puntos (x,y) se asocian en torno a  $\sim 1000$  configuraciones distintas obtenidas de la dinámica QM/MM MD y a partir de ellas se calcula la energía libre en ese punto que representa un conjunto de estructuras. EL mapa de energía para el estado S1 fue calculado con un  $\Delta SCF$  en el cual un electrón en el orbital HOMO (*Highest Occupied Molecular Orbital*) se pasa al orbital LUMO (*Lowest Unoccupied Molecular Orbital*) y se calcula la estructura electrónica de forma auto-consistente, ver sección anterior. Los perfiles de las barreras de energía en las figuras 5.4, 5.5 fueron calculados usando el programa MEPSA<sup>131</sup>.

En las simulaciones no adiabáticas el sistema es excitado al estado S1 en el primer paso y se deja que el sistema evolucione libremente (sin ningún

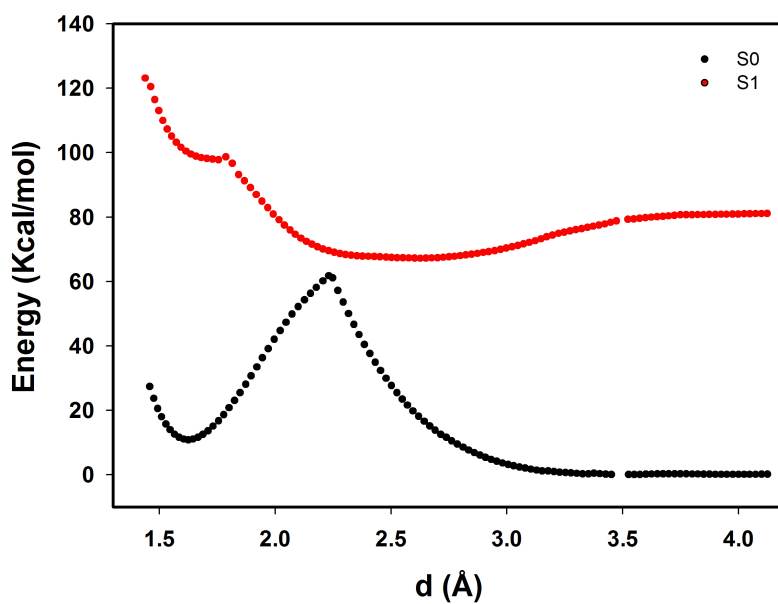
tipo de restricciones) durante 1500 fs (3000 pasos en cada trayectoria), usando la técnica de dinámica molecular con transiciones electrónicas explicada anteriormente e implementada en nuestro método de QM/MM. En estas trayectorias se empieza en la PES S1 pero en cada paso de dinámica hay una probabilidad de saltar al PES S0. La probabilidad de que suceda una transición está determinada por los acoplamientos no adiabáticos y la evolución de los estados electrónicos, usando el algoritmo '*fewest switches*'<sup>87</sup>. Esta probabilidad es despreciable excepto en el momento en el que el acople S1/S0 es grande, es decir en la CIZ.

Las posiciones y velocidades iniciales de estas trayectorias se eligieron de dinámicas QM/MM MD en el estado S0 en la región  $2.5 \text{ \AA} < d < 3.2 \text{ \AA}$  y  $75^\circ < \phi < 100^\circ$  (ver figuras 5.4 y 5.5). Estas configuraciones iniciales representan una fracción muy pequeña del espacio conformacional accesible a temperatura ambiente. Esta zona, cómo se vera posteriormente, corresponde a la región en la que una dinámica libre iría hacia la CIZ. De esta forma vamos a poder analizar la dinámica asociada con la transición  $S1 \rightarrow S0$  y, en particular, determinar cuantas de estas trayectorias forman el CTD y cuantas a través de una conversión interna vuelve a la configuración estándar del DNA.

### 5.2.3. Resultados: estudio adiabático

Para comparar nuestro método DFT, FIREBALL con los resultados obtenidos anteriormente se analizo la dimerización de timina en fase gaseosa y haciendo minimizaciones. En estos cálculos solo estaban las dos bases y se fue obligando con un potencial externo a acercarse la distancia 'd' De esta forma obtuvimos dos curvas de energía potencial para el estado basal y el estado excitado, ver figura 5.3. Los resultados obtenidos se asemejan mucho a tanto en geometría como energicamente a los obtenidos anteriormente<sup>83,116–122</sup>.





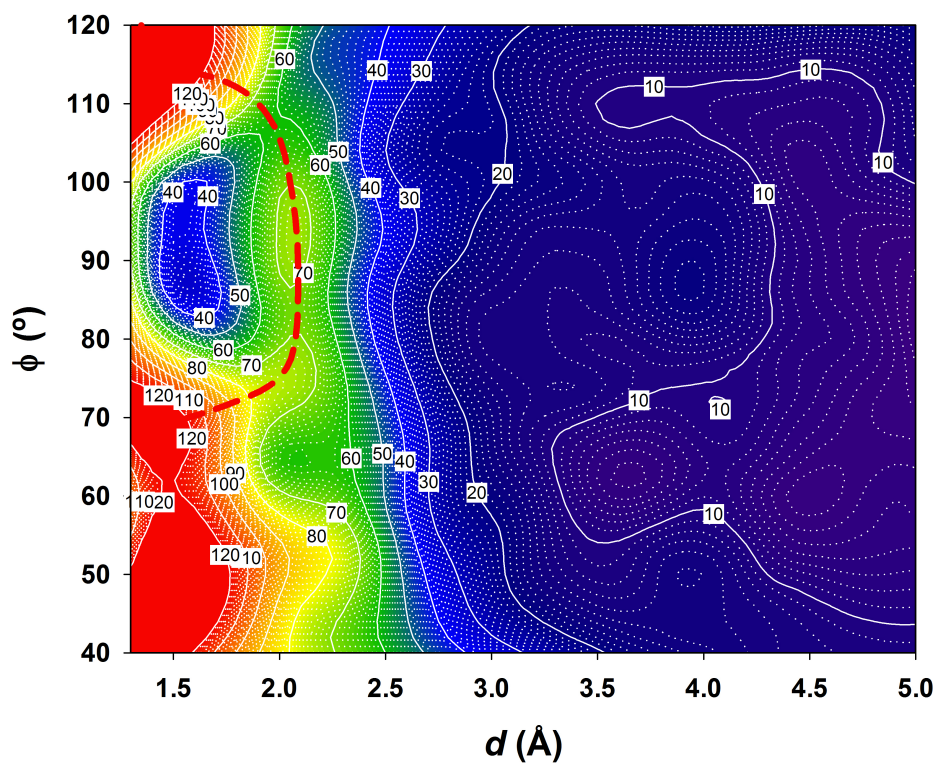
**Figura 5.3** Energía potencial del estado basal (S0) y el primer estado excitado (S1) para la reacción de formación del CTD entre dos bases de timina adyacentes en fase gaseosa. Como coordenada de reacción se ha utilizado la distancia  $d$  entre los puntos medios del enlace C5-C6.

En la figura 5.4 aparece el mapa de energía libre del estado fundamental en función de las coordenadas de reacción  $d$  y  $\phi$ . Como se aprecia en la figura 5.4 el ADN tiene un mínimo muy amplio en una región amplia definida por una distancia  $d$  están entorno a  $\sim 4 - 5 \text{ \AA}$  y par el ángulo  $\phi \sim 50^\circ - 100^\circ$ . Esta estructura el el estado normal del ADN. También se ve un mínimo en para los valores  $\phi \sim 90^\circ$  y  $d \sim 1.6 \text{ \AA}$  que corresponde a la fotolesión CTD (figura 5.2b) con una diferencia de energía de 33 kcal/min respecto al estado canónico. Entre estos dos mínimos de DNA estándar y CTD hay una barrera de alta energía (línea discontinua en la Fig. 5.4) de 60 - 70 kcal/mol, la cual es muy improbable de pasar a RT.

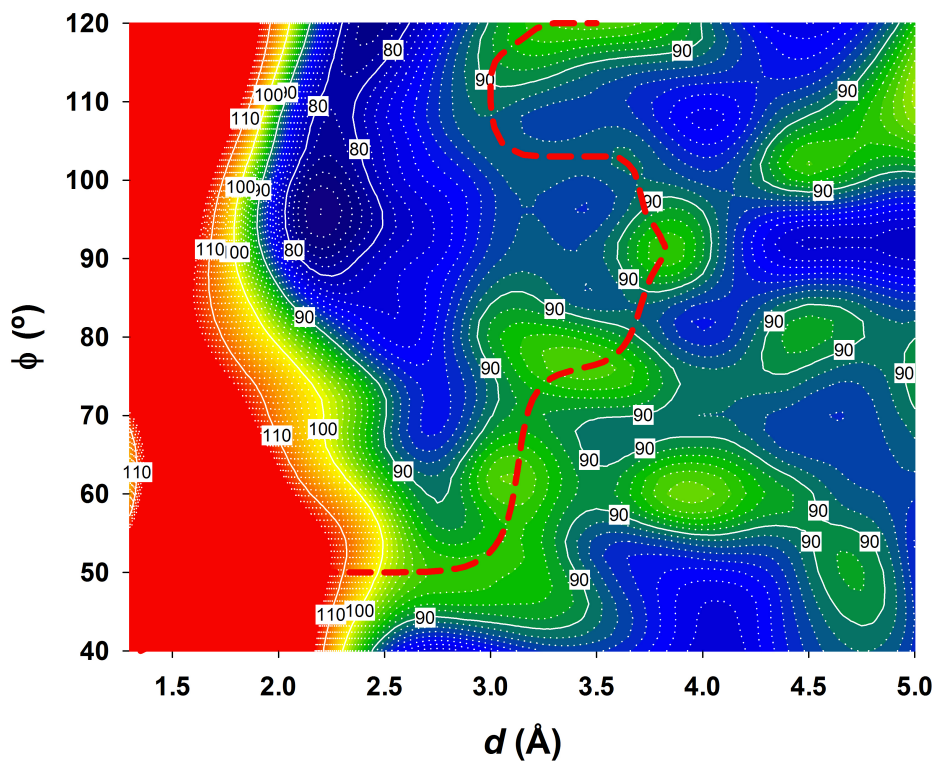
La figura 5.5 muestra la energía libre del primer estado excitado singlete S1. Ahora la región correspondiente a CTD (  $d \sim 1.6 \text{ \AA}$  y  $\phi \sim 90^\circ$  ) es claramente inestable; en este estado S1 las dos bases de timina no pueden acercarse demasiado a RT. Pero la característica principal es el mínimo de energía libre en  $d \sim 2.2 \text{ \AA}$  y  $\phi \sim 95^\circ$ ; este mínimo, como se discutirá posteriormente, es la zona de mayor probabilidad para que suceda la des-excitación de S1 a S0 y corresponde a la CIZ.

En este mapa hay variaciones respecto a los cálculos en fase gaseosa para dos timinas<sup>83,116-122</sup>, ya que en el ADN en forma biomolecular hay una barrera energética entre las configuraciones del estado canónico de ADN y la CTD (línea discontinua en la Fig. 5.5). Esta barrera define de forma precisa una región de conformaciones reactivas: configuraciones por encima y a la izquierda de la línea discontinua va a ir mayoritariamente a la CIZ pudiendo dar lugar a la formación del CTD, mientras que las que estén por debajo o a la derecha se moverán en dirección contraria hacia la configuración estándar del ADN.

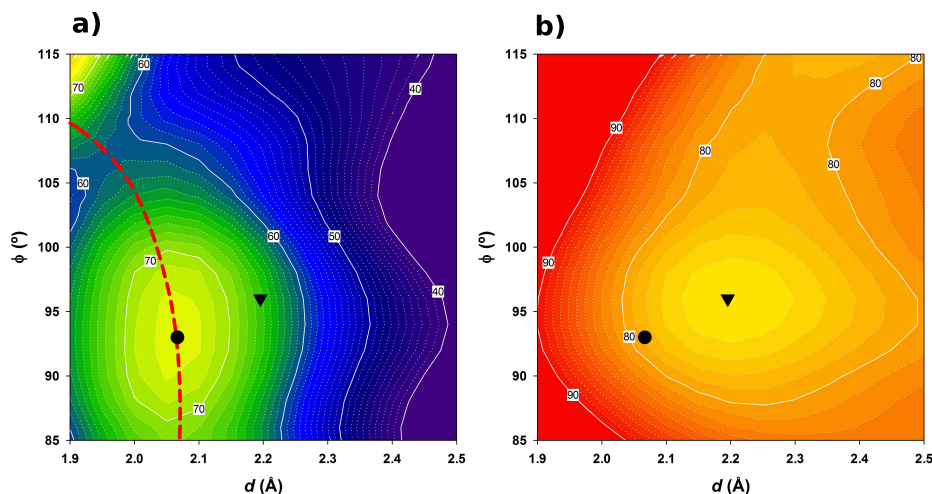
La zona de intersección cónica, CIZ, para el CTD es bastante amplia cómo se puede apreciar en la imagen del zoom de los mapas 5.6a y 5.6b donde podemos ver como el mínimo de S1 esta localizado cerca del máximo local de S0, Además en la región entre estos dos puntos ( $d \sim$



**Figura 5.4** Mapa de energía libre en kcal/mol en función de las coordenadas de reacción  $d$  y  $\phi$ , en el estado basal electrónico,  $S_0$ . La línea roja discontinua indica la barrera entre la conformación estándar del DNA y el CTD. Se ha utilizado el mínimo de energía como cero.



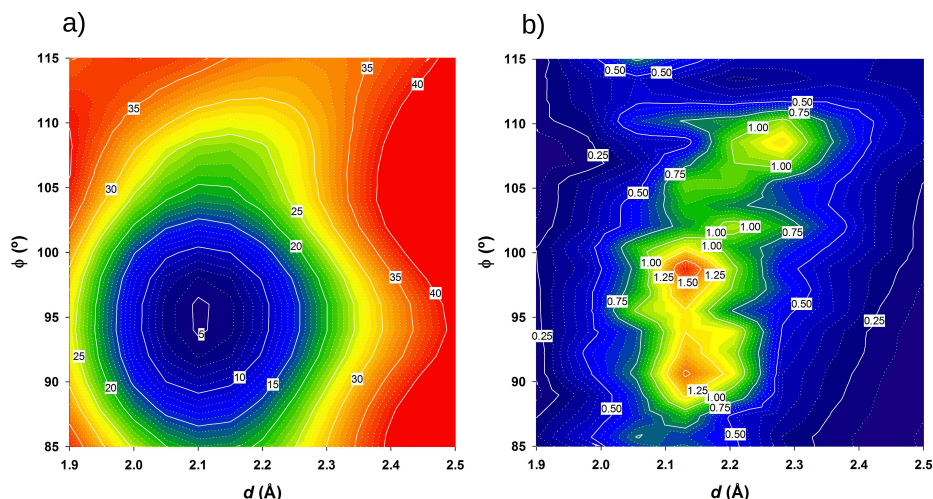
**Figura 5.5** Mapa de energía libre en kcal/mol en función de las coordenadas de reacción  $d$  y  $\phi$ , en el primer estado excitado, S1. La línea roja discontinua indica la barrera entre la región reactiva y la no reactiva del estado excitado S1. Se ha utilizado el mínimo de energía de S0 como cero.



**Figura 5.6** Zona de intersección cónica. Zoom de los mapas de energía libre en kcal/mol (fig. 5.4 y 5.5) alrededor de la zona de intersección cónica: (a) Estado S0; (b) estado S1. La línea roja discontinua en (a) indica la barrera entre las configuraciones del estado canónico del ADN y las configuraciones CTD, ver fig. 5.4.

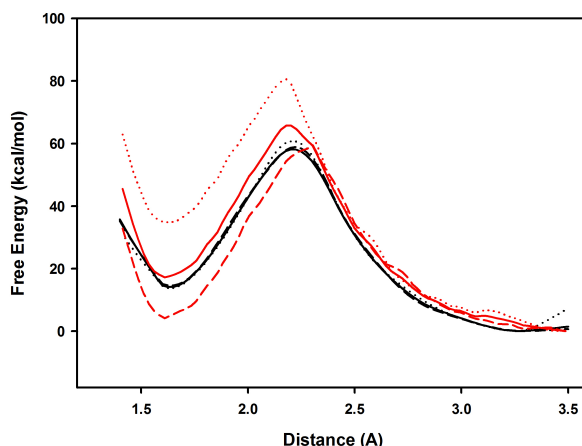
2.1 - 2.2 Å y  $\phi \sim 92^\circ - 98^\circ$ ), la energía libre de S1 y S0 tienen valores muy parecidos, ver imagen 5.7a. La geométrica característica en la CIZ en ambos estado es muy parecida y está definida por una distancia C6-C6',  $d_6$ ,  $\sim 1.7$  Å, mientras que la distancia C5-C5',  $d_5$ , todavía es bastante larga  $\sim 2.5$  Å, debido al efecto estérico de ambos grupos metilos.

Como la energía libre de las figuras 5.4, 5.5 y 5.6(a-b) es una cantidad estadística, cada punto del mapa está asociado a un conjunto de diferentes estructuras atómicas accesibles a RT. De esta cantidad de estructuras hay muchas de ellas en las que la diferencia de energía en la región CIZ de la imagen 5.6 entre los orbitales HOMO y LUMO es prácticamente cero y que por tanto tienen unos acoplamientos no adiabáticos (NACV) altos por lo que la probabilidad de transición  $S1 \rightarrow S0$  se vuelve muy alta, ver imagen 5.7b.



**Figura 5.7** Zona de intersección cónica. (a) Zoom de la diferencia de energía libre (fig. 5.4 y 5.5) alrededor de la zona de intersección cónica; (b) Mapa de la media del módulo del acoplamiento no adiabático para las diferentes geometrías de cada punto de la grid.

Para comparar el resultado al calcular la energía libre con el método propuesto en el capítulo 4 con métodos mas establecidos como WHAM<sup>56</sup> se realizó un estudio en vacío de la formación del CTD, ver figura 5.8. Las dinámicas se realizaron con las dos bases de timina y todo el sistema en QM. Para calcular la energía libre la distancia  $d$  entre 1.35 Å a 3.5 Å se hicieron 85 ventanas con el potencial fijo en una posición cada 0.025 Å. En cada ventana se utilizaron 1000 estructuras diferentes para calcular la energía libre en ambos métodos. Se puede apreciar como WHAM en este tipo de reacciones con barreras altas si se tiene solo en cuenta la reacción de ida o vuelta comete bastante error y hace falta combinar las dos. En cambio el método expuesto en el capítulo 4 tiene resultados mas parecidos en ambos casos haciendolo bastante idoneo para estudiar esta reacción.



**Figura 5.8** Perfiles de energía libre con el método expuesto en el capítulo 4 (líneas negras) y con WHAM (líneas rojas). Con dinámicas en varias direcciones, formación del CTD (línea punteada), del CTD al estado canónico (línea discontinua) y utilizando ambas dinámicas para obtener el perfil de energía libre (línea continuas).

#### 5.2.4. Resultados: estudio no adiabático

Para obtener mas información sobre el mecanismo atómico de esta reacción fotoquímica realizamos dinámicas moleculares QM/MM no adiabáticas para un conjunto de 2560 trayectorias usando la técnica de dinámica molecular con transiciones electrónicas explicada anteriormente. El interés de estas trayectorias es ver como el estado excitado ( $S_1$ ) evoluciona hasta llegar a la CIZ, y la posibilidad de formar el CTD después de la transición  $S_1 \rightarrow S_0$ . Para eso elegimos las coordenadas iniciales de estas trayectorias de forma que en el momento de la foto-excitación el sistema estuviese dentro de la región de conformaciones reactivas.

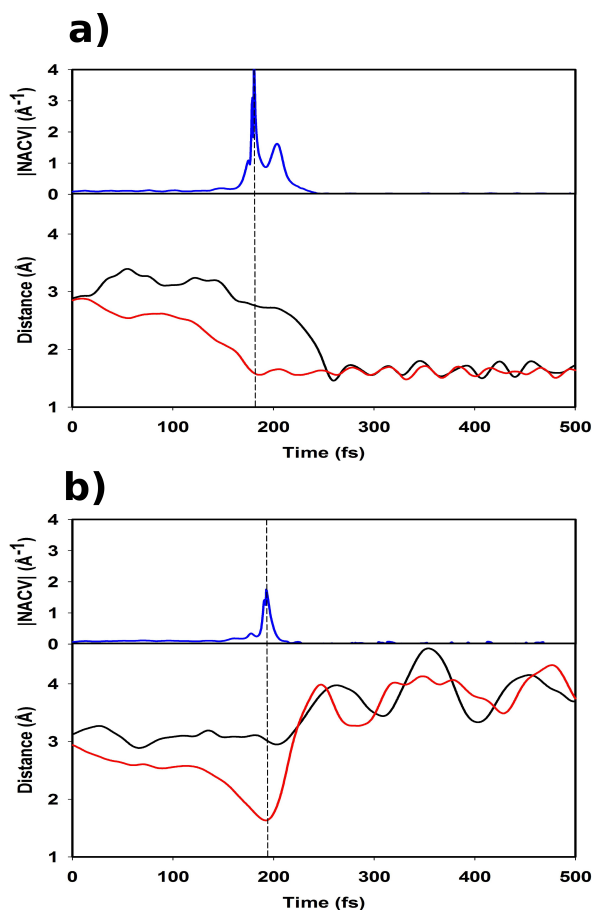
En las figuras 5.9(a-b) se puede ver dos ejemplos de estas trayectorias. Se puede apreciar que después de la foto-excitación inicial ( $t = 0$ ) el sistema se dirige rápidamente hacia la CIZ ( $\sim 200$  fs). Este proceso ultra-rápido esta dirigido por la atracción entre los átomos C6 de la timinas adyacentes: se observa como la distancia entre los dos átomos C6,  $d_{66}$ , se reduce

rápida hasta  $\sim 1.6 \text{ \AA}$ , creándose un incipiente enlace covalente entre las dos bases. En la CIZ el acoplamiento no adiabático entre los estados S1 y S0 es suficientemente grande, ver imagen 5.7b, y el sistema decae al estado S0 (líneas verticales discontinua en fig. 5.9). En ese momento el átomo C5 puede formar un segundo enlace covalente y formando la foto-lesión CTD (fig. 5.9a), o puede volver a la configuración estándar del DNA, rompiendo el recién formado enlace C6-C6' (fig. 5.9b).

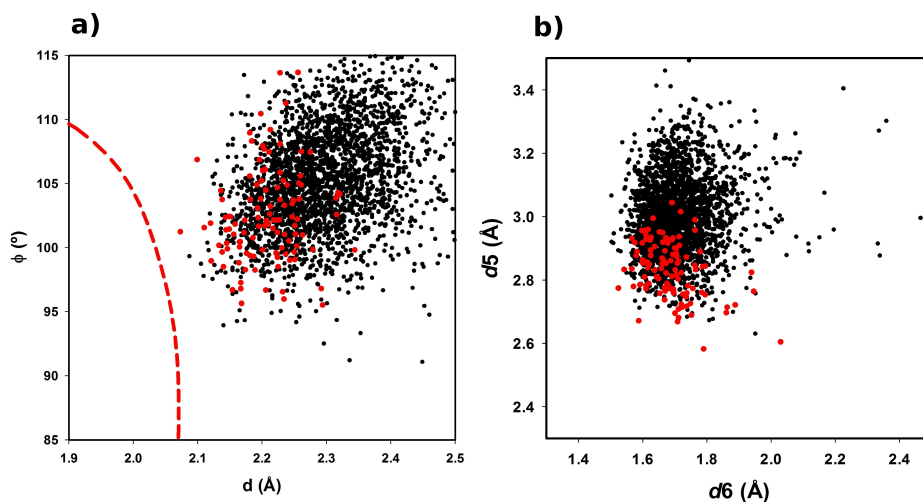
De forma inesperada solo 110 ( $\sim 4\%$ ) de las 2560 trayectorias terminaron formando la foto-lesión, mientras que las otras 2450 volvieron a la forma nativa del ADN. El valor de las coordenadas  $d$  y  $\phi$ , y de las distancias  $d5$  y  $d6$  en el momento de la transición se puede ver en las figuras 5.10a y 5.10b donde los puntos rojos corresponden a las que han terminado formando el CTD. En la figura 5.10a se puede observar que prácticamente todas las transiciones se producen en la zona que lleva a la estructura estándar del ADN 8a la derecha de la barrera que separa las regiones en el mapa de energía libre, valores altos de  $d$ ). Esto se debe a la topografía 'sloped'<sup>132</sup> de la CIZ (ver figuras 5.6(a-b)) y al ser un proceso ultra-rápido que está fuera del equilibrio. Esta zona de transición S1/S0 cubre una región bastante extensa en nuestro espacio conformacional,  $d \sim 2.1 - 2.5 \text{ \AA}$  y  $\phi \sim 90^\circ - 115^\circ$ , y se puede apreciar que hay una gran cantidad de estructuras diferentes con un alto valor de acoplamiento no adiabático, ver figura 5.7. En las figuras 5.10a y 5.10b también se puede apreciar como la formación del CTD es más probable para valores bajos de  $d$  y  $d5$ , es decir más cerca de la barrera que separa la conformación canónica de DNA y el CTD. También se aprecia como la transición suele producirse en valores más altos de  $d$  que los puntos de menor diferencia de energía de los mapas, ver figuras 5.6(a-b), esto refleja la naturaleza fuera del equilibrio de este proceso de foto-excitación al ser un proceso aleatorio.

También analizamos la evolución de los orbitales frontera (HOMO y LUMO) en la trayectoria no adiabática, en la figura 5.11 se puede ver un





**Figura 5.9** Ejemplo de las trayectorias no adiabáticas QM/MM: evolución temporal del módulo del NACV entre los estados S1 y S0 (línea azul), de la distancia C5-C5' (d5; línea negra) y de la distancia C6-C6' (d6; línea roja). La trayectoria (a) termina en la formación del CTD mientras que en la trayectoria (b) el sistema vuelve a la configuración estándar del DNA. Las líneas verticales discontinuas indican el momento en el que se produce la transición S1 → S0.



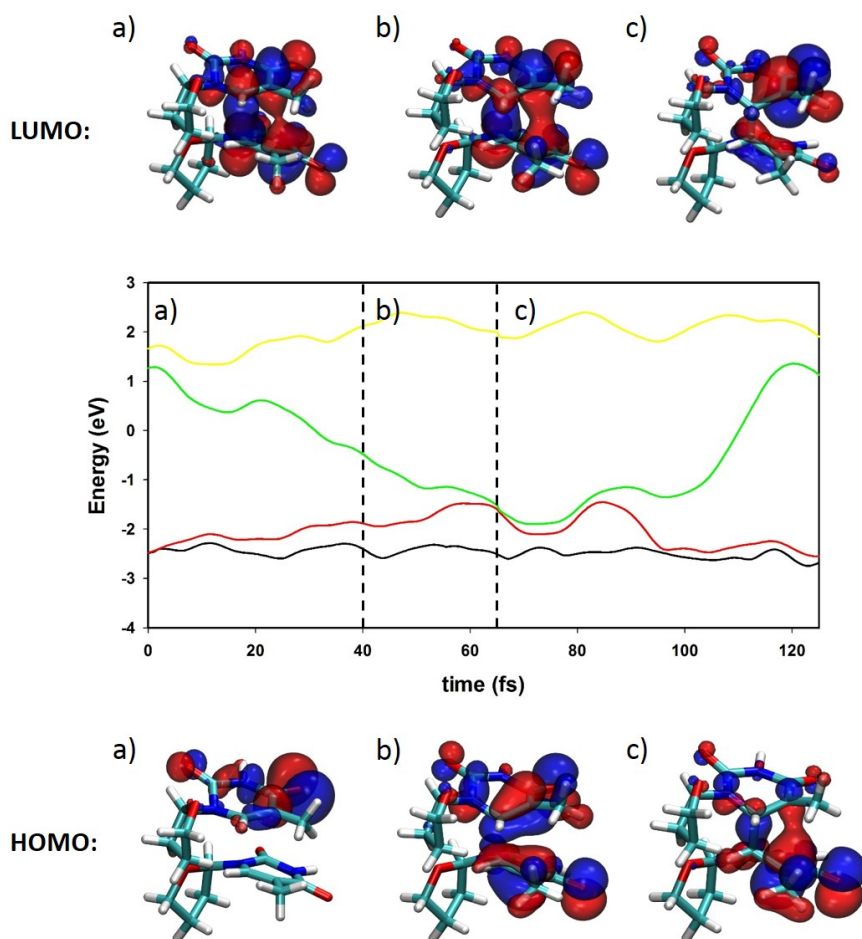
**Figura 5.10** Puntos correspondientes a las posiciones donde se produce las transiciones  $S1 \rightarrow S0$  en las simulaciones de las dinámicas no adiabáticas QM/MM; Los puntos rojos corresponden a las trayectorias que han terminado formando el CTD, solo una pequeña fracción de las trayectorias ( $\sim 4\%$ ) consiguen terminar formando el CTD. (a) En función de las coordenadas de reacción  $d$  y  $\phi$ . (b) En función de las distancias de los carbonos C6-C6',  $d6$ , y de los C5-C5',  $d5$ . La línea roja discontinua en (a) indica la barrera entre el estado canónico de DNA y las configuraciones CTD, ver fig. 5.4.

ejemplo de la evolución de estos orbitales en una trayectoria que termina formando el CTD. En ella se observa que el orbital LUMO tiene un carácter enlazantes en los estados a) y b) que promueve la formación de los enlace C-C, sobre todo el C6-C6. Esta es una característica general de todos los orbitales analizados que van a la CIZ. El orbital HOMO se aprecia que tiene un carácter claramente anti-enlazante en el estado b) pero justo en el momento de la transición  $S1 \rightarrow S0$  el orbital HOMO se vuelve enlazante y el LUMO anti-enlazante provocando la formación del CTD.

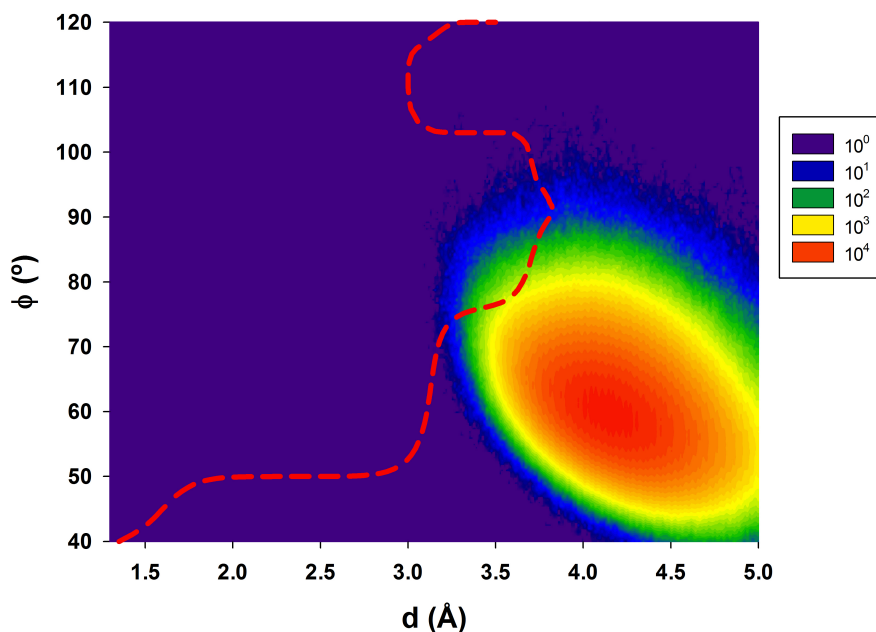
La probabilidad de que se produzca la formación del CTD una vez absorbido un fotón va a depender de dos factores, la probabilidad de llegar a la CIZ y la probabilidad de formar el CTD al producirse la desexcitación en la CIZ. La probabilidad de que al ser absorbido un fotón el sistema este dentro de la región activa y vaya a la zona de intersección cónica esta delimitada por la barrera calculada en la imagen 5.5. Para calcular la probabilidad de estar en la región se realizó una simulación larga de dinámica clásica (60 ns) a RT del estado fundamental  $S0$  y se analizó qué porcentaje de estados entraban en la región activa. En total se obtuvieron  $3 \cdot 10^7$  configuraciones distintas ( $\Delta t = 2 fs$ ) de las cuales  $\sim 1,1 \cdot 10^5$  alcanzaron la región activa, ver figura 5.12. Aunque el campo de fuerzas de la dinámica clásica no sea demasiado preciso para estados cercanos a la barrera de  $S1$  nos sirve para hacernos una idea del orden de magnitud de la fracción de tiempo,  $f \sim 10^{-2} - 10^{-3}$ , que el sistema pasa en la región activa.

### 5.2.5. Discusión y conclusiones

La formación de foto-lesiones en el ADN es un proceso muy complejo y hay que tener en cuenta una gran cantidad de factores. La conformación en el estado basal en el momento de la excitación, la probabilidad de estados excitados colectivos, el efecto de las bases adyacentes, estados excitados de vida larga, y muchos otros factores.<sup>107,125,133,134</sup> También un estudio mas detallado de la interacción de la radiación UV con el ADN (para estudiar



**Figura 5.11** Ejemplo de los orbitales frontera a lo largo de una trayectoria no adiabática. El panel de en media muestra la evolución temporal a lo largo de la trayectoria de los autovalores Kohn-Sham para los orbitales HOMO-1, HOMO, LUMO y LUMO+1. Arriba las forma de los orbitales HOMO en los momentos indicados a), b) y c) y abajo los orbitales HOMO. La transición  $S1 \rightarrow S0$  ocurre en el instante c).



**Figura 5.12** Distribución de las configuraciones del ADN en una dinámica clásica en función de las coordenadas de reacción para el estado basal. Los diferentes colores indican el numero de configuraciones, en escala logarítmica, para cada punto de la grid, en este caso se ha utilizado una grid mas fina con  $\sim 48000$  puntos y un  $\Delta d = 0.0125 \text{ \AA}$  y  $\Delta \phi = 0.5^\circ$ . La linea roja discontinua indica la barrera de la zona reactiva calculada en la figura 5.5

cambios en la probabilidad de absorción en función de la conformación) podría arrojar más información sobre el proceso.

Estos resultados sugieren el siguiente escenario para la formación del CTD en un ADN de doble hélice. Inicialmente un fotón UV es absorbido por las bases de timina moviendo la superficie de energía potencial de  $S_0$  a  $S_1$ . Si esta foto-excitación sucede cuando la geometría de las bases está en la zona reactiva el sistema irá rápidamente a la zona de intersección cónica a través de la PES  $S_1$ . Una vez que el sistema ha llegado a la CIZ existen una gran cantidad de geometrías donde hay un acoplamiento muy grande de los estados  $S_0$  y  $S_1$  y por lo tanto una gran posibilidad de volver al estado fundamental,  $S_0$ . Tras volver al estado  $S_0$  en la CIZ hay dos posibilidades: (a) volver a la forma a la configuración estándar del ADN (conversión interna) convirtiendo la energía del fotón absorbido en calor o (b) la formación de la foto-lesión CTD en el ADN. Debido a la topografía de la zona de intersección cónica la probabilidad de volver a la configuración inicial de las timinas es mucho mayor que la formación del CTD.

Estos resultados indican que la foto-estabilidad del ADN *vs.* formación del CTD radica en dos puntos distintos. El primero es la baja probabilidad al ser excitado de estar en la región activa y el segundo la baja probabilidad de que al producirse la des-excitación en la región de intersección cónica se forme el CTD. La baja probabilidad para el primer punto es debido a la estructura de doble hélice del DNA que crea la barrera entre estas zonas que no existe para el caso de dos timinas en vacío, ver figura 5.3, y también a que gracias a la geometría de la doble hélice la distancia de equilibrio de la coordenada  $d$  está mas alejada y la cantidad de geometrías reactivas es mucho menor en el estado basal. Ambos casos se deben a que la formación de los puentes de Watson-Crick y el 'backbone' del DNA estabilizan las dos timinas en una región mas estable y alejadas entre ellas. En el segundo punto la baja probabilidad de formación del CTD se debe a la geometría

'slope' entre las dos PES, S0 y S1, y a la naturaleza fuera del equilibrio del proceso de des-excitación en la CIZ.





## CAPÍTULO 6

### CONCLUSIONES GENERALES

---

En esta tesis se ha desarrollado un método de simulación QM/MM para poder simular reacciones enzimáticas con un gran balance entre eficiencia computacional y precisión. Para ello se ha mezclado el código de DFT FIREBALL para simular la región de QM y en la región clásica se ha utilizado el programa AMBER. Esta combinación permite estudiar de forma precisa y con dinámica sistemas complejos como son las biomoléculas. En esta implementación se ha tenido especial cuidado en como las cargas parciales de la región clásica eran capaces de modificar la estructura electrónica de la región QM y con ella conseguir simular el ambiente químico del entorno de la región activa. El efecto del ambiente sobre la región QM se ha comprobado sobre un puente salino, donde se ha visto la necesidad de tener en cuenta el ambiente sobre la estructura electrónica para poder modelar de forma correcta el sistema.

Para poder hacer de forma precisa los cálculos QM/MM se ha desarrollado una base de orbitales localizados FIREBALL ajustada para el uso de sistemas biológicos. En esta base se han tenido en cuenta diferentes bases de datos para biomoléculas y se ha conseguido una base con una gran precisión, comparada con métodos mucho más caros computacionalmente, y que está preparada para simular la gran mayoría de reacciones en biología al tener H, C, N, O, S, P y Mg.

Posteriormente se ha utilizado el método para simular dos reacciones enzimáticas y entender los distintos factores que afecta a las reacciones. Estas reacciones se han estudiado con superficies de energía libre en las cuales se pueden determinar el camino mas probable de la reacción. Para ello se ha utilizado en nuevo enfoque para explorar el espacio conformacional. La primera reacción estudiada la RNasa A. proteína encargada de la despolimerización de ARN. Para estudiarla se ha calculado una superficie de energía libre que nos ha dado información sobre las distintas trayectorias posibles y hemos podido discriminar el camino mas favorable enérgicamente. También hemos visto que residuos de la proteína son los que hacen posibles la catalización de la reacción. La otra enzima estudiada es la triosa fosfato isomerasa (TIM) esta enzima es parte esencial del ciclo de la glucólisis. En esta reacción gracias al estudio dinámico con los mapas de energía libre se ha propuesto un nuevo mecanismo de reacción donde hay una transferencia simultanea de dos protones. Esta transferencia simultanea ha sido posible de ver gracias a trabajar con dinámicas en lugar de minimizaciones.

Por último se ha estudiado una foto-reacción en el ADN, la dimerización del ciclobutano en timina. En esta reacción se ha localizado una zona de intersección cónica gracias al estudio de los mapas de energía libre del estado fundamental y primer estado excitado. También se ha visto que región al excitarse llevaría a la zona de intersección cónica y como la estructura del ADN protegla dimerización al evitar estar en esa zona. Después se ha estudiado con trayectorias no adiabáticas como se produce la desexcitación al llegar a la región de intersección cónica y la baja probabilidad que existe de llegar a forma el dímero de timina.

## PUBLICATIONS

---

- **Jesús I. Mendieta-Moreno**, Ross C. Walker, James P. Lewis, Paulino Gómez-Puertas, Jesús Mendieta y José Ortega. FIREBALL/AMBER: an efficient local-orbital DFT QM/MM method for biomolecular systems. *Journal of Chemical Theory and Computation*, 10, 2185–2193, 2014.
- **Jesús I. Mendieta-Moreno**, Iñigo Marcos-Alcalde, Daniel G. Trabada, Paulino Gómez-Puertas, José Ortega y Jesús Mendieta. Practical Quantum Mechanics Molecular Mechanics Method for the Dynamical Study of Reactions in Biomolecules. *Advances in protein chemistry and structural biology*, 100, 67–88, 2015.
- Vladimír Zobač, James P. Lewis, Enrique Abad, **Jesús I. Mendieta-Moreno**, Prokop Hapala, Pavel Jelínek y José Ortega. Photo-induced reactions from efficient molecular dynamics with electronic transitions using the FIREBALL local-orbital density functional theory formalism. *Journal of Physics: Condensed Matter*, 27, 175002, 2015.
- **Jesús I. Mendieta-Moreno**, Daniel G. Trabada, Jesús Mendieta, James P. Lewis, Paulino Gómez-Puertas y José Ortega. Quantum Mechanics/Molecular Mechanics Free Energy Maps and Nonadiabatic Simulations for a Photochemical Reaction in DNA:

Cyclobutane Thymine Dimer. *The Journal of Physical Chemistry Letters*, 7, 4391–4397, 2016.

Otras Publicaciones realizadas durante esta tesis:

- Fernando Martín-García, **Jesús I. Mendieta-Moreno**, Eduardo López-Viñas, Paulino Gómez-Puertas y Jesús Mendieta. The role of Gln 61 in HRas GTP hydrolysis: a quantum mechanics/molecular mechanics study. *Biophysical journal*, 102, 152–157, 2012.
- Fernando Martín-García, **Jesús I. Mendieta-Moreno**, Jesús Mendieta y Paulino Gómez-Puertas. Molecular dynamics analysis of conformational change of paramyxovirus F protein during the initial steps of membrane fusion. *Biochemical and biophysical research communications*, 420, 42–47, 2012.
- Fernando Martín-García, Estefanía Salvarelli, **Jesús I. Mendieta-Moreno**, Miguel Vicente, Jesús Mendieta y Paulino Gómez-Puertas. Molecular dynamics simulation of GTPase activity in polymers of the cell division protein FtsZ. *FEBS letters*, 586, 1236–1239, 2012.
- Fernando Martín-García, **Jesús I. Mendieta-Moreno**, Iñigo Marcos-Alcalde, Paulino Gómez-Puertas y Jesús Mendieta. Simulation of Catalytic Water Activation in Mitochondrial F1-ATPase Using a Hybrid Quantum Mechanics/Molecular Mechanics Approach: An Alternative Role for  $\beta$ -Glu 188. *Biochemistry*, 52, 959–966, 2013.
- Iñigo Marcos-Alcalde, Javier Setoain, **Jesús I. Mendieta-Moreno**, Jesús Mendieta y Paulino Gómez-Puertas. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics*, 31, 3853–3855, 2015.

- J. G. Vilhena, A. C. Dumitru, Elena T. Herruzo, **Jesús I. Mendieta-Moreno**, Ricardo Garcia, P. A. Serena y Rubén Pérez. Adsorption orientations and immunological recognition of antibodies on graphene. *Nanoscale*, 8, 13463–13475, 2016.
- Carmén M. Domínguez, Daniel Ramos, **Jesús Mendieta-Moreno**, José Fierro, Jesús Mendieta, Javier Tamayo y Montserrat Calleja. Effect of water-DNA interactions on elastic properties of DNA self-assembled monolayers. *Scientific Reports*, 7, 536, 2017.



# TABLAS Y GRÁFICAS DE LA OPTIMIZACIÓN DE LA BASE

---

Conjunto de enlaces utilizados para las distancias intra-moleculares (COV) y comparación de los resultados de fireball con los experimentales:

| Moléculas | distancia experimental (Å) | distancia en fireball (Å) |
|-----------|----------------------------|---------------------------|
| C-C       |                            |                           |
| HC-CH     | 1.20                       | 1.19369                   |
| CH2-CH2   | 1.33                       | 1.33225                   |
| CH3-CH3   | 1.54                       | 1.55024                   |
| Benzene   | 1.39                       | 1.40453                   |
| C-O       |                            |                           |
| CH2-OH    | 1.43                       | 1.43535                   |
| CH2-O     | 1.22                       | 1.28966                   |
| acetato   | 1.36                       | 1.35216                   |
| acetona   | 1.23                       | 1.28690                   |
| furano    | 1.43                       | 1.43984                   |
| peróxido  | 1.49                       | 1.53530                   |

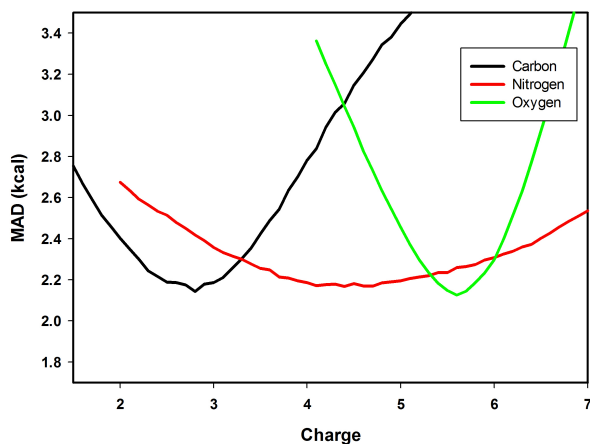
**Cuadro 1** Tabla de las distancias de enlace (COV).

| Moléculas                        | distancia experimental (Å) | distancia en fireball (Å) |
|----------------------------------|----------------------------|---------------------------|
| C-N                              |                            |                           |
| CH <sub>3</sub> -NH <sub>2</sub> | 1.47                       | 1.46235                   |
| CH <sub>3</sub> CH-NOH           | 1.28                       | 1.30851                   |
| CH <sub>3</sub> C-N              | 1.16                       | 1.17592                   |
| Guanidinio                       | 1.35                       | 1.37218512043             |
| C-H                              |                            |                           |
| CH <sub>3</sub> CH <sub>3</sub>  | 1.07                       | 1.06199                   |
| O-H                              |                            |                           |
| CH <sub>3</sub> O-H              | 0.96                       | 1.00683                   |
| H <sub>2</sub> O                 | 0.96                       | 1.00283                   |
| N-H                              |                            |                           |
| amoníaco                         | 1.017                      | 1.01392                   |

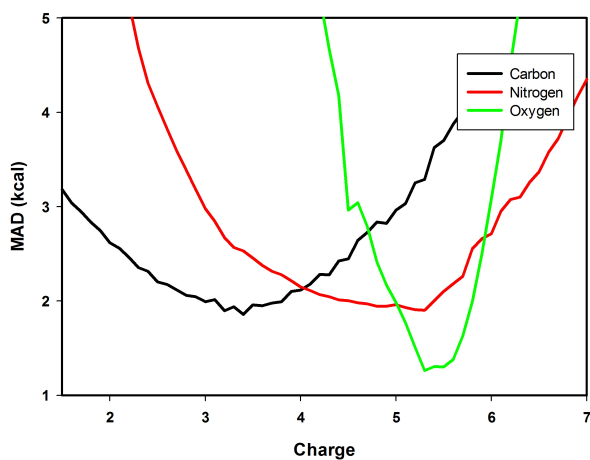
**Cuadro 2** Tabla de las distancias de enlace (COV).



Comparación de la variación del MAD con la geometría fija o con la geometría minimizada:



**Figura 1** Graficas de la variación del MAD en para el S66 con las estructuras fijas.



**Figura 2** Graficas de la variación del MAD en para el S66 con las estructuras mnimizadas.

Tablas de las energías de interacción para el S66 con la geometría fijada y minimizada para los parámetros seleccionados de la base:

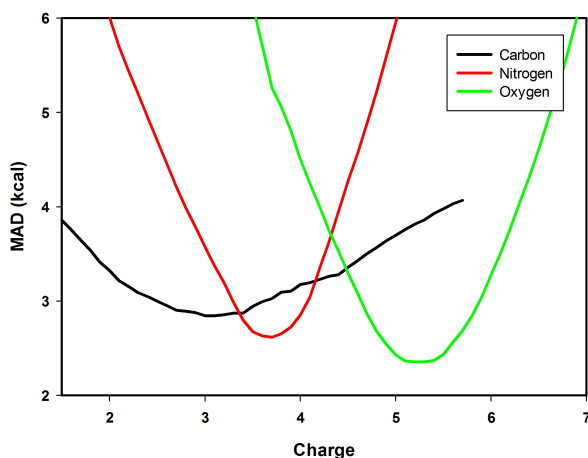
| Moléculas                                   | $\Delta E$ MP2 (Kcal) | $\Delta E$ S66 fix (Kcal) | $\Delta E$ S66 geo (Kcal) |
|---|-----------------------|---------------------------|---------------------------|
| water-water                                 | -4.50823              | -5.40752                  | -5.86852                  |
| water-MeOH                                  | -5.09395              | -5.81792                  | -6.89356                  |
| water-MeNH <sub>2</sub>                     | -6.38993              | -9.81415                  | -10.9006                  |
| water-peptide                               | -7.25929              | -8.12971                  | -9.95583                  |
| MeOH-MeOH                                   | -5.1493               | -4.36567                  | -5.35472                  |
| MeOH-MeNH <sub>2</sub>                      | -6.77964              | -7.18875                  | -7.90762                  |
| MeOH-peptide                                | -7.19472              | -6.61944                  | -8.18651                  |
| MeOH-water                                  | -4.51515              | -3.51658                  | -4.05386                  |
| MeNH <sub>2</sub> – <i>MeOH</i>             | -2.50432              | -1.20664                  | -2.07293                  |
| MeNH <sub>2</sub> – <i>MeNH<sub>2</sub></i> | -3.47284              | -0.56955                  | -1.66263                  |
| MeNH <sub>2</sub> – <i>peptide</i>          | -4.45289              | -2.14377                  | -2.93856                  |
| MeNH <sub>2</sub> – <i>water</i>            | -6.70354              | -7.90594                  | -10.2066                  |
| peptide-MeOH                                | -5.4583               | -5.03079                  | -6.00891                  |
| peptide-MeNH <sub>2</sub>                   | -6.61361              | -7.66256                  | -8.58143                  |
| peptide-peptide                             | -7.51064              | -7.17708                  | -8.0095                   |
| peptide-water                               | -4.60969              | -4.55336                  | -4.9178                   |
| uracil-uracil (BP)                          | -15.2796              | -18.6415                  | -19.5714                  |
| water-pyridine                              | -6.38301              | -6.68904                  | -6.90204                  |
| MeOH-pyridine                               | -6.78195              | -4.4518                   | -4.49236                  |
| AcOH-AcOH                                   | -17.3088              | -16.955                   | -17.3253                  |
| AcNH <sub>2</sub> – <i>AcNH<sub>2</sub></i> | -14.7054              | -18.8552                  | -19.9734                  |
| AcOH-uracil                                 | -17.6248              | -19.9342                  | -21.1826                  |
| AcNH <sub>2</sub> – <i>uracil</i>           | -17.3919              | -21.7811                  | -22.8029                  |
| benzene-benzene ( $\pi - \pi$ )             | -2.90556              | 1.58226                   | 1.54304                   |
| pyridine-pyridine ( $\pi - \pi$ )           | -4.04934              | 0.66761                   | 0.665973                  |
| uracil-uracil ( $\pi - \pi$ )               | -8.25087              | -2.58041                  | -6.21073                  |
| benzene-pyridine ( $\pi - \pi$ )            | -3.53971              | 1.21812                   | 1.19068                   |
| benzene-uracil ( $\pi - \pi$ )              | -5.1124               | -0.09844                  | -0.2231                   |
| pyridine-uracil ( $\pi - \pi$ )             | -6.20545              | -0.50240                  | -1.8665                   |
| benzene-ethene                              | -1.27983              | 1.55434                   | 1.56552                   |

**Cuadro 3** Tablas de las energías de interacción para el S66.

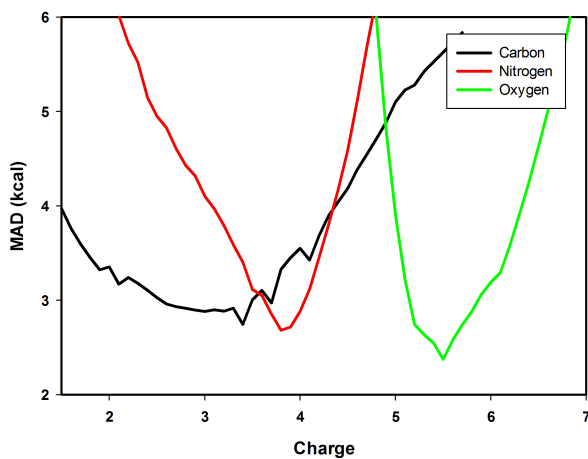
| Moléculas                              | $\Delta E$ MP2 (Kcal) | $\Delta E$ S66 fix (Kcal) | $\Delta E$ S66 geo (Kcal) |
|--|-----------------------|---------------------------|---------------------------|
| uracil-ethene                          | -2.76951              | 0.03514                   | -0.3099                   |
| uracil-ethyne                          | -3.25607              | -0.41030                  | -1.2639                   |
| pyridine-ethene                        | -1.71105              | 0.76010                   | 0.80228                   |
| pentane-pentane                        | -2.34751              | -0.79204                  | -1.03237                  |
| neopentane-pentane                     | -1.57039              | -1.79686                  | -1.78201                  |
| opentaneneopentane-nep                 | -0.993886             | -1.05347                  | -1.06812                  |
| cyclopentane-neopentane                | -1.42511              | -2.00297                  | -1.83419                  |
| cyclopentane-cyclopentane              | -1.86094              | -1.45792                  | -1.68158                  |
| benzene-cyclopentane                   | -2.95629              | -0.50817                  | -0.46154                  |
| benzene-neopentane                     | -2.34751              | -0.71061                  | -2.09228                  |
| uracil-pentane                         | -3.43363              | -1.24215                  | -1.80661                  |
| uracil-cyclopentane                    | -2.94015              | -0.11693                  | -0.66592                  |
| uracil-neopentane                      | -2.65421              | -0.81637                  | -1.15247                  |
| ethene-ethene                          | -1.27983              | -0.42508                  | -0.34257                  |
| ethyne-pentane                         | -1.33979              | 0.05034                   | 0.00368                   |
| peptide-pentane                        | -2.85483              | -0.32025                  | -1.17996                  |
| benzene-benze (TS)                     | -2.5758               | -0.57098                  | -0.58496                  |
| pyridine-pyridine (TS)                 | -3.17997              | -0.77698                  | -0.78159                  |
| benzene-pyridine (TS)                  | -2.9978               | -0.75053                  | -0.74887                  |
| benzene-ethyne (CH- $\pi$ )            | -2.64729              | -1.04909                  | -0.96888                  |
| ethyne-ethyne (TS)                     | -1.33979              | -1.13335                  | -0.99649                  |
| benzene-AcOH (OH- $\pi$ )              | -4.14619              | -4.22459                  | -4.50528                  |
| benzene-AcNH <sub>2</sub> (NH- $\pi$ ) | -3.75186              | -1.44662                  | -1.54359                  |
| benzene-water (OH- $\pi$ )             | -2.88019              | -5.35827                  | -5.93235                  |
| benzene-MeOH (OH- $\pi$ )              | -3.63426              | -4.28695                  | -4.46165                  |
| benzene-MeNH <sub>2</sub> (NH- $\pi$ ) | -2.73953              | -1.914                    | -3.26403                  |
| benzene-peptide (NH- $\pi$ )           | -4.67426              | -2.83562                  | -2.90916                  |
| pyridine-pyridine (CH-N)               | -3.57661              | -2.00871                  | -2.01228                  |
| ethyne-water (CH-O)                    | -2.55274              | -1.85707                  | -1.79215                  |
| ethyne-AcOH (OH- $\pi$ )               | -4.26379              | -5.82726                  | -6.16567                  |
| pentane-AcOH                           | -1.86555              | -1.05765                  | -1.34652                  |
| pentane-AcNH <sub>2</sub>              | -2.31984              | -1.36398                  | -1.55563                  |
| benzene-AcOH                           | -3.20995              | 0.06565                   | -0.16584                  |
| peptide-ethene                         | -2.25757              | -0.78288                  | -1.15388                  |
| pyridine-ethyne                        | -3.67576              | -3.21212                  | -2.88308                  |
| NeNH <sub>2</sub> -pyridine            | -3.4198               | -0.83551                  | -1.0294                   |

**Cuadro 4** Tablas de las energías de interacción para el S66.

Comparación de la variación del MAD con la geometría fija o con la geometría minimizada:



**Figura 3** Gráficas de la variación del MAD en para el IHB con las estructuras fijas.



**Figura 4** Gráficas de la variación del MAD en para el IHB con las estructuras mnimizadas.

Tabla de las energías de interacción para el IHB con la geometría fijada y minimizada para los parámetros seleccionados de la base:

| Moléculas                   | $\Delta E$ MP2 (Kcal) | $\Delta E$ IHB fix (Kcal) | $\Delta E$ IHB geo (Kcal) |
|-----------------------------|-----------------------|---------------------------|---------------------------|
| acetate-methanol            | -19,74858             | -18,71544                 | -18,98301                 |
| acetate-water               | -21,05839             | -24,76906                 | -25,21938                 |
| acetate-methylamine         | -11,46082             | -8,164000                 | -7,585932                 |
| methylammonium-formaldehyde | -19,10059             | -23,28352                 | -23,30028                 |
| methylammonium-methylamine  | -28,55981             | -38,25852                 | -34,72591                 |
| methylammonium-methanol     | -21,22903             | -25,22678                 | -27,69838                 |
| methylammonium-water        | -18,51026             | -22,02677                 | -22,34004                 |
| guanidinium-formaldehyde    | -18,09057             | -19,57014                 | -19,60351                 |
| guanidinium-methylamine     | -20,20056             | -21,35192                 | -19,83701                 |
| guanidinium-methanol        | -19,79009             | -20,13435                 | -22,70817                 |
| guanidinium-water           | -17,47025             | -16,80218                 | -17,36399                 |
| imidazolium-formaldehyde    | -16,40949             | -16,89670                 | -16,45895                 |
| imidazolium-methylamine     | -25,97939             | -34,12416                 | -30,12881                 |
| imidazolium-methanol        | -18,9092              | -19,54823                 | -21,81146                 |
| imidazolium-water           | -16,49020             | -17,35938                 | -17,80817                 |

**Cuadro 5** Tablas de las energías de interacción para el IHB.

Para las mediciones de  $d$  inter se han utilizado las distancias de puente de hidrógeno que hay en S66 (las 23 primeras) y en IHB (todas). En la siguiente tabla se muestra los datos optimizados con MP2 y con la base optimizada con fireball:

| Moléculas                             | $d$ MP2 (Å) | $d$ fireball (Å) |
|---------------------------------------|-------------|------------------|
| water-water                           | 1.963       | 1.9634           |
| water-MeNH <sub>2</sub>               | 1.961       | 1.9614           |
| water-peptide                         | 1.861       | 1.8614           |
| MeOH-MeOH                             | 1.905       | 1.9058           |
| MeOH-MeNH <sub>2</sub>                | 1.938       | 1.9386           |
| MeOH-peptide                          | 1.855       | 1.8553           |
| MeOH-water                            | 1.952       | 1.9522           |
| MeNH <sub>2</sub> – MeOH              | 2.199       | 2.1996           |
| MeNH <sub>2</sub> – MeNH <sub>2</sub> | 2.244       | 2.2443           |
| MeNH <sub>2</sub> – peptide           | 2.203       | 2.2038           |
| MeNH <sub>2</sub> – water             | 1.941       | 1.9413           |
| peptide-MeOH                          | 1.989       | 1.9898           |
| peptide-MeNH <sub>2</sub>             | 2.048       | 2.0486           |
| peptide-peptide                       | 1.947       | 1.9470           |
| peptide-water                         | 2.052       | 2.0524           |
| uracil-uracil (BP)                    | 1.824       | 1.8244           |
| water-pyridine                        | 1.953       | 1.9538           |
| MeOH-pyridine                         | 1.931       | 1.9316           |
| AcOH-AcOH                             | 1.681       | 1.6813           |
| AcNH <sub>2</sub> – AcNH <sub>2</sub> | 1.844       | 2.8647           |
| AcOH-uracil                           | 1.767       | 1.7678           |
| AcNH <sub>2</sub> – uracil            | 1.730       | 1.7305           |

**Cuadro 6** Tablas de las energías de interacción para el IHB.

| Moléculas                   | $d$ MP2 (Å) | $d$ fireball (Å) |
|-----------------------------|-------------|------------------|
| acetate-methanol            | 1.692       | 1.5852           |
| acetate-water               | 2.042       | 1.9406           |
| acetate-methylamine         | 1.995       | 1.9716           |
| methylammonium-formaldehyde | 1.729       | 1.6593           |
| methylammonium-methylamine  | 1.106       | 1.2971           |
| methylammonium-methanol     | 1.642       | 1.5757           |
| methylammonium-water        | 1.705       | 1.6087           |
| guanidinium-formaldehyde    | 2.012       | 1.9743           |
| guanidinium-methylamine     | 1.787       | 1.5178           |
| guanidinium-methanol        | 1.977       | 1.9513           |
| guanidinium-water           | 2.034       | 2.0111           |
| imidazolium-formaldehyde    | 1.745       | 1.7241           |
| imidazolium-methylamine     | 1.596       | 1.3580           |
| imidazolium-methanol        | 1.639       | 1.6335           |
| imidazolium-water           | 1.700       | 1.6625           |

**Cuadro 7** Tablas de las energías de interacción para el IHB.





## BIBLIOGRAFÍA

---

- [1] J. D. Watson and F. H. C Crick. A Structure for Deoxyribose Nucleic Acid. *Nature*, 171:737–738, 1953. doi: 10.1038/171737a0.
- [2] E. Fermi, J. Pasta, and S. Ulam. Studies of Nonlinear problems. *Los Alamos report*, 1955.
- [3] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method. , 31:459–466, August 1959. doi: 10.1063/1.1730376.
- [4] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Physical Review*, 136:405–411, October 1964. doi: 10.1103/PhysRev.136.A405.
- [5] Andrew J Mccammon, Bruce R Gelin, and Martin Karplus. Dynamics of folded protein. *Nature*, 267:585–590, 1977. doi: 10.1038/267585a0.
- [6] A Warshel and M Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.
- [7] MJ Field, PA Bash, and M Karplus. A Combined Quantum-Mechanical And Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.*, 11:700–733, 1990.
- [8] Hans Martin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angew. Chem.*, 48:1198–1229, 2009.
- [9] Mo. Karplus and J.A. McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, 9:646–652, 2002.
- [10] M. K. Gilson and H.-X. Zhou. Calculation of Protein-Ligand Binding Affinities. *Ann. Rev. Biophys. Biomol. Struct.*, 36:21–42, 2007.

- [11] R. M. Martin. *Electronic Structure*. Cambridge University Press, 2004.
- [12] D. Marx and J. Hutter. *Ab initio Molecular Dynamics*. Cambridge University Press, 2009.
- [13] Fernando Martín-García, Jesús Ignacio Mendieta-Moreno, Eduardo López-Viñas, Paulino Gómez-Puertas, and Jesús Mendieta. The Role of Gln61 in HRas GTP hydrolysis: a quantum mechanics/molecular mechanics study. *Biophys. J.*, 102:152–157, 2012.
- [14] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim. Experimental Observation of the Quantum Hall Effect and Berry's Phase in Graphene. *Nature*, 438:201, 2005. doi: 10.1038/nature04235.
- [15] J.I. Mendieta-Moreno, R. Walker, J.P. Lewis, P. Gomez-Puertas, J. Mendieta, and J. Ortega. FIREBALL / AMBER : An Efficient Local-Orbital DFT QM / MM Method for Biomolecular Systems. *J. Chem. Theory Comput.*, 10:2185–2193, 2014.
- [16] J.P. Lewis, P. Jelinek, J. Ortega, A.A. Demkov, D.G. Trabada, B. Haycock, H. Wang, G. Adams, J.K. Tomfohr, E. Abad, H. Wang, and D.A. Drabold. Advances and applications in the fireball ab initio tight-binding molecular-dynamics formalism. *Phys. Stat. Sol. B*, 248:1989–2007, 2011.
- [17] Pavel Jelínek, Hao Wang, James Lewis, Otto Sankey, and José Ortega. Multicenter approach to the exchange-correlation interactions in ab initio tight-binding methods. *Physical Review B*, 71:235101, 2005.
- [18] J.P. Lewis, K.R. Glaesemann, G.A. Voth, J. Fritsch, A.A. Demkov, J. Ortega, and O.F. Sankey. Further developments in the local-orbital density-functional-theory tight-binding method. *Phys. Rev. B*, 64: 195103, 2001.
- [19] Alexander A. Demkov, José Ortega, Otto F. Sankey, and Matthew P. Grumbach. Electronic structure approach for complex silicas. *Phys. Rev. B*, 52:1618–1630, 1995.
- [20] Otto F. Sankey and David J. Niklewski. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys. Rev. B*, 40:3979–3995, 1989.

- [21] D.A. Case, T.A. Darden, III. Cheatham, T.E., C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. *AMBER 12*, 2012.
- [22] R. Salomon-Ferrer, D. Case, and R. Walker. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.*, 3:198–210, 2012.
- [23] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136, 1964.
- [24] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.
- [25] M. A. Basanta, Y.J. Dappe, P. Jelínek, and J. Ortega. Optimized atomic-like orbitals for first-principles tight-binding molecular dynamics. *Computational Materials Science*, 39:759 – 766, 2007.
- [26] John P Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048, 1981.
- [27] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098, 1988.
- [28] C. Lee, W. Yang, and R. G. Parr. . *Phys. Rev. B*, 37:785, 1988.
- [29] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77:3865, 1996.
- [30] Axel D. Becke. A New Mixing of Hartree-Fock and Local Density Functional Theories. *J. Chem. Phys.*, 98:1372, 1993.
- [31] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for Mixing Exact Exchange with Density Functional Approximations. *J. Chem. Phys.*, 105:9982, 1996.
- [32] J. Harris. Simplified method for calculating the energy of weakly interacting fragments. *Phys. Rev. B*, 31:1770, 1985.

- [33] W. Matthew C. Foulkes and Roger Haydock. Tight-binding models and density-functional theory. *Phys. Rev. B*, 39:12520–12536, 1989.
- [34] F. J. García-Vidal, J. Merino, R. Pérez, R. Rincón, J. Ortega, and F. Flores. Density-functional approach to LCAO methods. *Phys. Rev. B*, 50:10537–10547, 1994.
- [35] K Schönhammer, O Gunnarsson, and RM Noack. Density-functional theory on a lattice: Comparison with exact numerical results for a model with strongly correlated electrons. *Phys. Rev. B*, 52:2504, 1995.
- [36] This approximation is related to the use of a selfconsistent version of the Harris-Foulkes functional.<sup>19,32,33</sup>
- [37] Andrew P. Horsfield. Efficient *ab initio* tight binding. *Phys. Rev. B*, 56:6594–6602, 1997.
- [38] P Löwdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.*, 18:365, 1950.
- [39] B. C. Carlson and Joseph M. Keller. Orthogonalization procedures and the localization of wannier functions. *Phys. Rev.*, 105:102–103, 1957.
- [40] RC Walker, MF Crowley, and DA Case. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.*, 29:1019–1031, 2008.
- [41] M. Kusic, T. Matamoros, M. Nevot, J. Mendieta, J. Martinez-Picado, M.A. Martínez, and L Menéndez-Arias. Thymidine analogue excision and discrimination modulated by mutational complexes including single amino acid deletions of asp-67 or thr-69 in hiv-1 reverse transcriptase. *J. Biol. Chem.*, 286:20615–20624, 2011.
- [42] Martin Korth and Walter Thiel. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods Are Almost As Accurate and Robust As DFT-GGA Methods for Organic Molecules. *Journal of Chemical Theory and Computation*, 7:2929–2936, 2011.
- [43] Stewart a Adcock and J Andrew Mccammon. Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins Molecular

- Dynamics : Survey of Methods for Simulating the Activity of Proteins. 106(5):1589–1615, 2006.
- [44] Bill R Miller III, T Dwight McGee Jr, Jason M Swails, Nadine Homeyer, Holger Gohlke, and Adrian E Roitberg. Mmpbsa. py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation*, 8(9):3314–3321, 2012.
- [45] Ulf Ryde and Par Soderhjelm. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chemical Reviews*, 116(9):5520–5566, 2016. ISSN 15206890. doi: 10.1021/acs.chemrev.5b00630.
- [46] Donald G Truhlar, Bruce C Garrett, and Stephen J Klippenstein. Current status of transition-state theory. *The Journal of physical chemistry*, 100(31):12771–12800, 1996.
- [47] Judith P Klinman and Amnon Kohen. Hydrogen tunneling links protein dynamics to enzyme catalysis. *Annual review of biochemistry*, 82:471–496, 2013.
- [48] Donald G Truhlar and Bruce C Garrett. Variational transition-state theory. *Accounts of Chemical Research*, 13(12):440–448, 1980.
- [49] Donald G Truhlar, Jiali Gao, Mireia Garcia-Viloca, Cristobal Alhambra, Jose Corchado, Maria Luz Sanchez, and Tina D Poulsen. Ensemble-averaged variational transition state theory with optimized multidimensional tunneling for enzyme kinetics and other condensed-phase reactions. *International journal of quantum chemistry*, 100(6):1136–1152, 2004.
- [50] H Bernhard Schlegel. Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry*, 3(2): 214–218, 1982.
- [51] Marc W van der Kamp and Adrian J Mulholland. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52(16):2708–28, April 2013. ISSN 1520-4995. doi: 10.1021/bi400215w.
- [52] Matthew G Quesne, Tomasz Borowski, and Sam P de Visser. Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs. *Chemistry–A European Journal*, 22(8): 2562–2581, 2016.

- [53] Gideon J. Davies, Antoni Planas, and Rovira Carme. Conformational analyses of the reaction coordinate of glycosidases. *Accounts of chemical research*, 45(2):308–316, 2012. ISSN 15204898. doi: 10.1021/ar2001765.
- [54] Shina CL Kamerlin, Maciej Haranczyk, and Arie Warshel. Progress in ab initio qm/mm free-energy simulations of electrostatic energies in proteins: accelerated qm/mm studies of p k a, redox reactions and solvation free energies. *The Journal of Physical Chemistry B*, 113(5):1253–1272, 2008.
- [55] Sanghyun Park, Fatemeh Khalili-Araghi, Emad Tajkhorshid, and Klaus Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *The Journal of Chemical Physics*, 119:3559, 2003.
- [56] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.
- [57] Miguel Jorge, Nuno M. Garrido, António J. Queimada, Ioannis G. Economou, and Eugénia a. MacEdo. Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration. *Journal of Chemical Theory and Computation*, 6(4):1018–1027, 2010.
- [58] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562–12566, 2002.
- [59] Robert W. Zwanzig. High-Temperature Equation of State by a Perturbation, Method. I. Nonpolar Gases. *The Journal of Chemical Physics*, 22(8):1420–1426, 1954.
- [60] James J P Stewart. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *Journal of molecular modeling*, 10(2):155–164, April 2004.
- [61] William S Cleveland and Susan J Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fittings. *Journal of the American Statistical Association*, 83(403):596–610, 1986.

- [62] Dorothy Findlay, D. G. Herries, A. P. Mathias, B. R. Rabin, and Ross C. A. © 1961 Nat. *Nature*, 190:781–784, 1961.
- [63] Claudi M Cuchillo, M Victòria Nogués, and Ronald T Raines. Bovine pancreatic ribonuclease: fifty years of the first enzymatic reaction mechanism. *Biochemistry*, 50(37):7835–7841, September 2011. ISSN 1520-4995. doi: 10.1021/bi201075b.
- [64] I Zegers, D Maes, M H Dao-Thi, F Poortmans, R Palmer, and L Wyns. The structures of RNase A complexed with 3'-CMP and d(CpA): active site conformation and conserved water molecules. *Protein science : a publication of the Protein Society*, 3(12):2322–2339, December 1994. ISSN 0961-8368. doi: 10.1002/pro.5560031217.
- [65] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins*, 65(May):712–725, 2006.
- [66] Timothy M Glennon and Arieh Warshel. Energetics of the Catalytic Reaction of Ribonuclease A : A Computational Study of Alternative Mechanisms. *Journal of the American Chemical Society*, 7863(98): 10234–10247, 1998.
- [67] Michael J. S. Dewar, Eve G. Zebisch, Eamonn F. Healy, and James J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 4: 3902–3909, 1985.
- [68] Petra Imhof, Frank Noé, Stefan Fisher, and Jeremy C. Smith. AM1/d Parameters for Magnesium in Metalloenzymes. *Journal of Chemical Theory and Computation*, 4:1050–1056, 2006.
- [69] James J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modelling*, 13:1173–1213, 2006.
- [70] Yang Yang, Haibo Yu, Darrin York, Marcus Elstner, and Quiang Cui. Description of Phosphate Hydrolysis Reactions with the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) Theory. 1. Parameterization. *Journal of Chemical Theory and Computation*, 4:2067–2084, 2008.

- [71] Alan Hall and Jeremy R Knowles. The uncatalyzed rates of enolization of dihydroxyacetone phosphate and of glyceraldehyde 3-phosphate in neutral aqueous solution. The quantitative assessment of the effectiveness of an enzyme catalyst. *Biochemistry*, 14(19):4348–4353, 1975.
- [72] P a Bash, M J Field, R C Davenport, G a Petsko, D Ringe, and M Karplus. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, 30(24): 5826–5832, 1991.
- [73] P J Lodi and J R Knowles. Neutral imidazole is the electrophile in the reaction catalyzed by triosephosphate isomerase: structural origins and catalytic implications. *Biochemistry*, 30(28):6948–6956, 1991.
- [74] Harold Walba and Robert W Isensee. Acidity Constants of Some Arylimidazoles and Their Cations. *The Journal of organic chemistry*, 26(8):2789–2791, 1961.
- [75] Qiang Cui and Martin Karplus. Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: Verification of methodology and analysis of reaction mechanisms. *Journal of Physical Chemistry B*, 106(7):1768–1798, 2002.
- [76] Maybelle K. Go, Tina L. Amyes, and John P. Richard. Rescue of K12G triosephosphate isomerase by ammonium cations: The reaction of an enzyme in pieces. *Journal of the American Chemical Society*, 132 (38):13525–13532, 2010.
- [77] Moumita Samanta, M. R N Murthy, Hemalatha Balaram, and Padmanabhan Balaram. Revisiting the Mechanism of the Triosephosphate Isomerase Reaction: The Role of the Fully Conserved Glutamic Acid 97 Residue. *ChemBioChem*, 12(12): 1886–1896, 2011.
- [78] W W Cleland and Maurice M Kreevoy. Low-Barrier Hydrogen Bonds. *Science*, 264(June):1887–1890, 1994.
- [79] Hiroshi Ishikita and Keisuke Saito. Proton transfer reactions and hydrogen-bond networks in Proton transfer reactions and hydrogen-bond networks in protein environments. *Journal of the Royal Society Interface*, 11(91), 2014.



- [80] Gerwald Jogl, Sharon Rozovsky, Ann E McDermott, and Liang Tong. Optimal alignment for enzymatic proton transfer: structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1):50–55, 2003.
- [81] Maybelle K. Go, Astrid Koudelka, Tina L. Amyes, and John P. Richard. Role of Lys-12 in catalysis by triosephosphate isomerase: A two-part substrate approach. *Biochemistry*, 49(25):5377–5389, 2010.
- [82] Hao Hu, Zhenyu Lu, and Weitao Yang. QM/MM Minimum Free Energy Path: Methodology and Application to Triosephosphate Isomerase. *J. Chem. Theory Comput.*, 3(2):390–406, 2007.
- [83] Shaila Rössle, Jana Friedrichs, and Irmgard Frank. The formation of DNA photodamage: The role of exciton localization. *ChemPhysChem*, 11(9):2011–2015, 2010. ISSN 14394235. doi: 10.1002/cphc.201000081.
- [84] Mario Barbatti, Jaroslaw J. Szymczak, Adlia J A Aquino, Dana Nachtigallov, and Hans Lischka. The decay mechanism of photoexcited guanine - A nonadiabatic dynamics study. *Journal of Chemical Physics*, 134(1), 2011. ISSN 00219606.
- [85] B. F. E. Curchod, U. Rothlisberger, and I. Tavernelli. Trajectory-based nonadiabatic dynamics with time-dependent density functional theory. *ChemPhysChem*, 14:1314–1340, 2013.
- [86] N. L. Doltsinis. Molecular dynamics beyond the born-oppenheimer approximation: mixed quantum classical approaches. *Computational Nanoscience: Do It Yourself*, 31:389–409, 2006.
- [87] J.C. Tully. Molecular dynamics with electronic transitions. *Journal of Chem. Phys.*, 93:93, 1990.
- [88] J.C. Tully and Preston R. K. Effects of surface crossing in chemical reactions: the H<sub>3</sub><sup>+</sup> system. *Journal of Chem. Phys.*, 55:562, 1971.
- [89] R. J. Maurer and K. Reuter. Assessing computationally efficient isomerization dynamics: SCF density-functional theory study of azobenzene molecular switching. *J. Chem. Phys.*, 135:224303, 2011.
- [90] E.K.U. Gross and W Kohn. Time-dependent density-functional theory. *Advances in quantum chemistry*, 21:255–291, 1990.

- [91] Miguel AL Marques and Eberhard KU Gross. Time-dependent density functional theory. *Annu. Rev. Phys. Chem.*, 55:427–455, 2004.
- [92] Colleen F. Craig, Walter R. Duncan, and Oleg V. Prezhdo. Trajectory Surface Hopping in the Time-Dependent Kohn-Sham Approach for Electron-Nuclear Dynamics. *Phys. Rev. Lett.*, 95:163001, 2005.
- [93] Enrique Abad, James P Lewis, Vladmír Zobač, Prokop Hapala, Pavel Jelínek, and José Ortega. Calculation of non-adiabatic coupling vectors in a local-orbital basis set. *The Journal of chemical physics*, 138(15):154106, 2013.
- [94] Sharon Hammes-Schiffer and John C Tully. Proton transfer in solution: Molecular dynamics with quantum transitions. *The Journal of chemical physics*, 101(6):4657–4667, 1994.
- [95] DF Coker and L Xiao. Methods for molecular dynamics with nonadiabatic transitions. *The Journal of chemical physics*, 102(1): 496–510, 1995.
- [96] Michael F Herman. Nonadiabatic semiclassical scattering. i. analysis of generalized surface hopping procedures. *The Journal of chemical physics*, 81(2):754–763, 1984.
- [97] Wolfgang J Schreier, Peter Gilch, and W. Zinth. Early events of DNA photodamage. *Annu. Rev. Phys. Chem.*, 66:497–519, 2015.
- [98] Jean Cadet, Stéphane Mouret, Jean Luc Ravanat, and Thierry Douki. Photoinduced damage to cellular DNA: Direct and photosensitized reactions. *Photochemistry and Photobiology*, 88(5):1048–1065, 2012. ISSN 00318655. doi: 10.1111/j.1751-1097.2012.01200.x.
- [99] Thierry Douki. Effect of denaturation on the photochemistry of pyrimidine bases in isolated DNA. *Journal of Photochemistry and Photobiology B: Biology*, 82(1):45–52, 2006. ISSN 10111344. doi: 10.1016/j.jphotobiol.2005.08.009.
- [100] Katarina Volkovova, Dagmar Bilanicova, Alena Bartonova, Silvia Letašiová, and Maria Dusinska. Associations between environmental factors and incidence of cutaneous melanoma. Review. *Environmental health : a global access science source*, 11 Suppl 1(Suppl 1):S12, 2012. ISSN 1476-069X. doi: 10.1186/1476-069X-11-S1-S12.

- [101] T. Lindhal and R.D. Wook. Quality control by DNA repair. *Science*, 286:1897, 1999.
- [102] Andrea Christa Kneuttinger, Gengo Kashiwazaki, Stefan Prill, Korbinian Heil, and Thomas Carell. Formation and direct repair of UV-induced dimeric DNA pyrimidine lesions. *Photochemistry and Photobiology*, 90(1):1–14, 2014. ISSN 00318655. doi: 10.1111/php.12197.
- [103] Dominik B. Beukers, Alexander Schlueter, Thomas Carell, and Wolfgang Zinth. Watson-Crick Base Pairing Controls Excited - State Decay in Natural DNA. *Angewandte Chemie - International Edition*, 53(42):11366–11369, 2014. ISSN 15213773. doi: 10.1002/anie.201406286.
- [104] Carlos E Crespo-Hernández, Boiko Cohen, Patrick M Hare, and Bern Kohler. Ultrafast excited-state dynamics in nucleic acids. *Chemical Reviews*, 104(4):1977–2019, apr 2004. ISSN 0009-2665. doi: 10.1021/cr0206770.
- [105] Sylvie Marguet and Dimitra Markovitsi. Time-resolved study of thymine dimer formation. *Journal of the American Chemical Society*, 127(16):5780–5781, 2005. ISSN 00027863. doi: 10.1021/ja050648h.
- [106] Wolfgang J Schreier, Tobias Schrader, F.O. Koller, Peter Gilch, C.E. Crespo-Hernandez, V. N. Swaminathan, T. Carell, W. Zinth, and B. Kohler. Thymine dimerization in DNA is an ultrafast photoreaction. *Science*, 315:625, 2007.
- [107] Wolfgang J Schreier, Julia Kubon, Nadja Regner, Karin Haiser, Tobias E Schrader, W. Zinth, Pascale Clivio, and Peter Gilch. Thymine dimerization in DNA model systems: cyclobutane photolesion is predominantly formed via the singlet channel. *Journal of the American Chemical Society*, 131(14):5038–9, apr 2009. ISSN 1520-5126. doi: 10.1021/ja900436t.
- [108] Chris T Middleton, Kimberly de La Harpe, Charlene Su, Yu Kay Law, Carlos E Crespo-Hernández, and Bern Kohler. DNA Excited-State Dynamics: From Single Bases to the Double Helix. *Annual Review of Physical Chemistry*, 60(1):217–239, 2009. ISSN 0066-426X. doi: 10.1146/annurev.physchem.59.032607.093719.

- [109] Thomas Gustavsson, Roberto Improta, and Dimitra Markovitsi. DNA/RNA: Building blocks of life under UV irradiation. *Journal of Physical Chemistry Letters*, 1(13):2025–2030, 2010. ISSN 19487185. doi: 10.1021/jz1004973.
- [110] J. Chen, Y. Zhang, and B. Kohler. Excited states in DNA strands investigated by ultrafast laser spectroscopy. *Top. Curr. Chem.*, 356:39, 2015.
- [111] J. Cadet, A. Grand, and T. Douki. Solar UV Radiation-Induced DNA Bipyrimidine Photoproducts: Formation and Mechanistic Insights. *Top. Curr. Chem.*, 356:249–276, 2015.
- [112] Yu Kay Law, Javad Azadi, Carlos E Crespo-Hernández, Eric Olmon, and Bern Kohler. Predicting thymine dimerization yields from molecular dynamics simulations. *Biophysical Journal*, 94(9): 3590–600, may 2008. ISSN 1542-0086. doi: 10.1529/biophysj.107.118612.
- [113] Andrew T. Johnson and Olaf Wiest. Structure and dynamics of poly(T) single-strand DNA: Implications toward CPD formation. *Journal of Physical Chemistry B*, 111(51):14398–14404, 2007. ISSN 15206106. doi: 10.1021/jp076371k.
- [114] Y. Lu, Z. Lan, and W. Thiel. Computational modeling of photoexcitation in DNA single and double strands. *Top. Curr. Chem.*, 356:89–122, 2015.
- [115] F. Plasser, A.J.A. Aquino, H. Lischka, and D. Nachtigallova. Electronic excitation processes in single-strand and double-strand DNA: a computational approach. *Top. Curr. Chem.*, 356:1–38, 2015.
- [116] Martial Boggio-Pasqua, Gerrit Groenhof, Lars V Schäfer, Helmut Grubmüller, and Michael A Robb. Ultrafast deactivation channel for thymine dimerization. *Journal of the American Chemical Society*, 129(36):10996–7, sep 2007. ISSN 0002-7863. doi: 10.1021/ja073628j.
- [117] L Blancafort and A Migani. Modeling Thymine Photodimerization DNA: Mech & Correlation Diagram. *J. Am. Chem. Soc.*, (47): 14540–14541. doi: 10.
- [118] Daniel Roca-Sanjuán, Gloria Olaso-González, Israel González-Ramírez, Luis Serrano-Andrés, and Manuela Merchán.

- Molecular basis of DNA photodimerization: Intrinsic production of cyclobutane cytosine dimers. *Journal of the American Chemical Society*, 130(32):10768–10779, 2008. ISSN 00027863. doi: 10.1021/ja803068n.
- [119] Juan J. Serrano-Pérez, I. González-Ramírez, Pedro B. Coto, M Merchán, and Luis Serrano-Andrés. Theoretical insight into the intrinsic ultrafast formation of cyclobutane pyrimidine dimers in UV-irradiated DNA: Thymine versus cytosine. *Journal of Physical Chemistry B*, 112(45):14096–14098, 2008.
- [120] Mario Barbatti. Computational Reference Data for the Photochemistry of Cyclobutane Pyrimidine Dimers. *ChemPhysChem*, 15(15):3342–3354, 2014.
- [121] Bo Durbeej and Leif A. Eriksson. Reaction mechanism of thymine dimer formation in DNA induced by UV light. *Journal of Photochemistry and Photobiology A: Chemistry*, 152(1-3):95–101, 2002. ISSN 10106030. doi: 10.1016/S1010-6030(02)00180-6.
- [122] R. B. Zhang and L. A. Eriksson. A triplet mechanism for the formation of cyclobutane pyrimidine dimers in UV-irradiated DNA. *Journal of Physical Chemistry B*, 110:7556, 2006.
- [123] David R. Yarkony. Diabolical conical intersections. *Reviews of Modern Physics*, 68(4):985–1013, 1996. ISSN 00346861. doi: 10.1103/RevModPhys.68.985.
- [124] S. Matsika and P. Krause. Nonadiabatic events and conical intersections. *Annu. Rev. Phys. Chem.*, 62:621–643, 2011. ISSN 0066-426X. doi: 10.1146/annurev-physchem-032210-103450.
- [125] Wook Lee and Spiridoula Matsika. QM/MM studies reveal pathways leading to the quenching of the formation of thymine dimer photoproduct by flanking bases. *Phys. Chem. Chem. Phys.*, 17: 9927–9935, 2015. ISSN 1463-9076. doi: 10.1039/C5CP00292C.
- [126] Jesús I. Mendieta-Moreno, I. Marcos-Alcalde, Daniel G. Trabada, Paulino Gómez-Puertas, José Ortega, and Jesús Mendieta. A Practical Quantum Mechanics Molecular Mechanics Method for the Dynamical Study of Reactions in Biomolecules. *Adv Protein Chem Struct Biol.*, 100:67–88, 2015.

- [127] Vladimír Zobač, James P Lewis, Enrique Abad, Jesús I Mendieta-Moreno, Prokop Hapala, Pavel Jelínek, and José Ortega. Photo-induced reactions from efficient molecular dynamics with electronic transitions using the FIREBALL local-orbital density functional theory formalism. *Journal of Physics: Condensed Matter*, 27(17):175002, 2015.
- [128] HaJeung Park, Kaijiang Zhang, Yingjie Ren, Sourena Nadji, Nanda Sinha, John-Stephen Taylor, and ChulHee Kang. Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25): 15965–70, dec 2002. ISSN 0027-8424. doi: 10.1073/pnas.242422699.
- [129] R. Salomon-Ferrer, David A. Case, and Ross C. Walker. An Overview of the Amber biomolecular simulation packages. *WIREs Comput. Mol. Sci.*, 3:198–210, 2013.
- [130] Martin Fuchs and Matthias Scheffler. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory. *Computer Physics Communications*, 119: 67 – 98, 1999.
- [131] I. Marcos-Alcalde, J. Setoain, Jesus I. Mendieta-Moreno, J. Mendieta, and P. Gómez-Puertas. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics*, 31:3853–3855, 2015.
- [132] Benjamin G Levine and Todd J Martínez. Isomerization through conical intersections. *Annual Review of Physical Chemistry*, 58: 613–634, 2007. ISSN 0066-426X. doi: 10.1146/annurev.physchem.57.032905.104612.
- [133] Dimitra Markovitsi. UV-induced DNA Damage: The Role of Electronic Excited States. *Photochemistry and Photobiology*, 92(1): 45–51, 2016. ISSN 17511097. doi: 10.1111/php.12533.
- [134] Ignacio Vayá, Johanna Brazard, Miquel Huix-Rotllant, Arun K. Thazhathveetil, Frederick D. Lewis, Thomas Gustavsson, Irene Burghardt, Roberto Improta, and Dimitra Markovitsi. High-Energy Long-Lived Mixed Frenkel-Charge-Transfer Excitons: From Double Stranded (AT) <sub>n</sub> to Natural DNA. *Chemistry - A European Journal*, 22(14):4904–4914, 2016. ISSN 09476539. doi: 10.1002/chem.201504007.